AD-A035 067 CACI INC-FEDERAL ARLINGTON VA F/G 6/7
STATISTICAL ANALYSES APPLIED TO THE U. S. NAVY AIRCREW AUTOMATE--ETC(U)
NOV 76 H L CROWSON N00019-76-C-0034
UNCLASSIFIED NAVWESA-R-7606 NL

1 OF 1
AD
A035067

1 (2)

R-7606

STATISTICAL ANALYSES APPLIED TO THE U.S. NAVY AIRCREW AUTOMATED ESCAPE SYSTEMS

Henry L. Crowson

CACI, Inc. - Federal

Naval Weapons Engineering Support Activity
Weapon Systems Analysis Department (ESA-19)
Washington Navy Yard
Washington, D.C. 20374

NOVEMBER 1976

Prepared for

Crew Systems Division (AIR-531)
NAVAL AIR SYSTEMS COMMAND
Jefferson Plaza 2
Washington, D.C. 20361

ACKNOWLEDGMENTS

The author of this report expresses his sincere appreciation to the many persons who assisted with its preparation. In particular, he acknowledges the support, constructive comments, and technical discussions from the following: Mr. Fred Guill (NAVAIR), and Mr. Joe Bizup and Ms. Susan Vaitekunas (NAVWESA-19). He is indebted to Steve Linefsky (CACI) for obtaining many computer results. He appreciates the artwork performed by Mary Kijanka. A special note of thanks is due his typist, Juanita Blanchard, for the typing, preparing all the tables, and managing the production of this document.

Approved

J. Watson

Department Head
Weapon Systems Analysis Department

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: 19 R-7606 18 NAVWESA	2. GOVT ACCESSION NO.
	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): 6 STATISTICAL ANALYSES APPLIED TO THE U.S. NAVY AIRCREW AUTOMATED ESCAPE SYSTEMS,	5. TYPE OF REPORT & PERIOD COVERED
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): 10 Henry L./Crowson	8. CONTRACT OR GRANT NUMBER(s): 15 N00019-76-C-0034 New
9. PERFORMING ORGANIZATION NAME AND ADDRESS: CACI, Inc. 1815 North Fort Myer Drive Arlington, Virginia 22209	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: Crew Systems Division (AIR-531) NAVAL AIR SYSTEMS COMMAND Jefferson Plaza 2 Washington, D. C. 20361	12. REPORT DATE: 11 30 November 1976
	13. NUMBER OF PAGES: 158 12 146p.
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): Naval Weapons Engineering Support Activity Weapon Systems Analysis Department (ESA-19) Washington Navy Yard, Washington, D.C. 20374	15. SECURITY CLASS. (of this report): UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Companion volume to "R-7605 Statistical Analysis Methodology for USN Aircrew Automated Escape Systems".

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Aircrew-automated escape system	Non-parametric analysis	Deterministic analysis
Ejection system	Hypothesis testing	
Analysis of variance	Fatality patterns	
Injury Analysis	Prediction	

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Statistical analyses methodologies, such as analysis of variance, non-parametric tests and hypothesis tests were applied to uncover trends and dependencies of ejection-related injuries. Ejection-related fatality patterns for the A-6 aircraft were developed deterministically. The A-4, A-5, A-6, A-7, F-4, F-8 and F-9 ejection-related injuries were investigated, and parent probability density functions derived. The A-4, A-6, A-7, F-4 and F-8 contain an underlying injury trend. Fatalities occur randomly upon ejection from the next page

DD FORM 1473, 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601 |

407 136

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

bpg

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

cont.

A-4, A-7, F-4 and F-8 aircraft. Fatalities do not occur randomly from the A-6 aircraft. Differences were revealed among various ejection fatality scenarios: hardware hazards, aircrew judgment, environmental conditions, aircraft associated with ejections, and ejection seats used in ejection.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF TABLES

LIST OF TABLES (Continued)

LIST OF ILLUSTRATIONS

Section 1.0

EXECUTIVE SUMMARY

1.0 Executive Summary

This is a report of the analysis of the U.S. Navy Aircrew Automated Escape Systems (AAES) project from 15 May 1976 to 15 August 1976. It addresses the following four main areas: (a) Brief survey of the analytical techniques used, (2) Reasons for using the given techniques, (3) Summary of results obtained, and (4) Suggestions for future investigation.

1.1 Analytical Techniques Used

Primary emphasis for this study was the application of well known statistical and deterministic methods to ejection related personnel injury data. A flow diagram of analyses used and a general sequence in which they were applied, as detailed in this report, is shown in Figure 2-3. Techniques used include the following:

- Two-Way Analysis of Variance
- Non-Parametric Trend Test
- Non-Parametric Run Test
- Deterministic Analyses
  - -- Numerical Techniques
  - -- Deterministic Prediction
- Hypothesis Testing using the following density functions:
  - -- Runs Discrete Density Function
  - -- Normal Density Function
  - -- Gamma Density Function
  - -- Exponential Density Function
- Chi-Square Density Function was used to test Goodness-of-Fit of random samples to the preceding parent probability density functions

Application of these techniques, as well as other techniques, are detailed in our Phase I Final Technical Report entitled, "Statistical Analysis Methodology for the U.S. Navy Aircrew Automated Escape Systems," dated 15 May 1976.

1.2 Rationale for Using the Statistical and Deterministic Techniques Employed

In any program of the kind addressed in this report, where a well-defined analytical path does not exist, inevitably some trial and error will result: Some techniques tried and shelved; other techniques, not originally spelled out, will be tried with excellent results. Experiments to be performed on the given data frequently is a major problem. Questions which are asked frequently dictate the analyses to be applied. Consider the following:

1.2.1. Does any relationship exist among the various high performance aircraft ejection related fatalities and causes of the fatalities?

This question can be answered by application of the Two-Way Analysis of Variance Matrix, One Observation per Cell.

1.2.2. Does any relationship exist between two separate ejection seats used and causes of ejection related fatalities?

This too is a question which can be answered by a Two-Way Analysis of Variance Technique.

1.2.3. Do injury data sets contain an underlying trend?

This can be answered by application of the non-parametric trend test.

1.2.4. On a chronological basis, do ejection related fatalities occur randomly over time?

This question can be answered upon application of the non-parametric run test.

1.2.5. Can future ejection related fatality patterns, as well as other patterns, be predicted mathematically?

Contingent upon results from hypothesis testing and non-parametric tests, this question can be answered by applying deterministic and numerical techniques.

1.2.6. Can future ejection related injury patterns be statistically predicted?

The answer to this question hinges upon proper formulation and testing of hypotheses about the ejection related injury data sets under investigation.

These questions are representative, not exhaustive. Results of our investigations are summarized below.

1.3 Results Obtained from Applying Statistical and Deterministic Analyses to Ejection Related Injury Data

In statistical analyses, answers are rarely absolute. Generally they are given with a confidence level attached. Subject to this understanding, the following are results obtained to date:

1.3.1(A) A difference exists among the average number of ejection related fatalities caused by (a) hardware hazards, (b) aircrew judgment, and (c) environmental conditions, when ejections are considered on an aircraft-by-aircraft basis.

1.3.1(B) A difference exists among the average number of ejection related fatalities associated with ejection from the A-4, A-7, A-6, F-4, and F-8 U.S. Navy high performance aircraft.

1.3.2(A) No difference exists among the average number of ejection related fatalities caused by (a) hardware hazards, (b) aircrew judgment, and (c) environmental conditions, when ejections are considered from ESCAPAC or Martin-Baker ejection seats.

1.3.2(B) No difference exists among the average number of ejection related fatalities upon ejection with either the ESCAPAC or Martin-Baker ejection seats.

1.3.3. On a chronological basis, the A-4, A-6, A-7, F-4, and F-8 ejection related injury data individually contain underlying trends.

1.3.4. On a chronological basis, fatalities occur randomly upon ejection from the A-4, A-7, F-4, and F-8 aircraft.

1.3.5. On a chronological basis, fatalities do not occur randomly upon ejection from the A-6 aircraft.

1.3.6. Based on the indicated non-random behavior in the A-6 data, future ejection related fatality patterns have been predicted.

1.3.7. If a parent probability density function can be derived, future ejection related injury patterns can be predicted. These injury patterns have been predicted in this report for the A-6 aircraft (generalized gamma probability density function); the A-4 aircraft (exponential density function); and seven aircraft (A-4, A-5, A-6, A-7, F-4, F-8, F-9) taken all together as a random sample of size 977 ejections (exponential density function).

1.3.8. For the A-6 aircraft, it was discovered that the least number of ejection related fatalities occurred when the aircraft was in the velocity range of 200 to 300 knots.

1.3.9. A comparison of the A-6 ejection related injury pattern and the injury pattern for seven aircraft (A-4, A-5, A-6, A-7, F-4, F-8, F-9) reveals that on a percentage basis, the A-6 injuries exceed those injuries from the seven aircraft studied, in every injury category except minimal injuries.

1.3.10. The seven aircraft mentioned above have a fatality rate, over the years studied (1969 through 1975), of 14.7 percent of all ejections. The A-6 fatality rate is 19.6 percent of all A-6 ejections. Thus, the A-6 percentage fatality rate is 33.34 percent higher than the fatality rate for the seven aircraft.

1.4 Suggestions for Future Investigations

The statistical investigations reported herein, while interesting and informative, should be viewed as beginning investigations only. They have highlighted a real need for in-depth study of various problem areas. A summary of areas which have an immediate need to be studied are the following:

1.4.1. An in-depth investigation of all A-6 ejections.

1.4.2. Determine the underlying injury trends in the A-4, A-6, A-7, F-4, and F-8 ejection related injury data.

1.4.3. Reconcile any apparently inconsistent results detected by the analysis of variance technique.

1.4.4. After sufficient analytical results have been obtained from both preliminary and in-depth investigations, the question of why particular ejection injury patterns exist needs to be addressed.

1.4.5. Statistical analysis techniques should continue to be applied to additional ejection data.

Section 2.0

INTRODUCTION

2.0 Introduction

This report contains detailed descriptions of results obtained upon applying statistical analyses to actual Medical Officer's Reports (MORs) ejection injury data. These data were developed on a case-by-case study of each aircrew ejection from U.S. Navy high performance aircraft over the time period 1 January 1969, through 31 December 1975. Statistical analyses applied were developed in CACI's report, "Statistical Analysis Methodology for the U.S. Navy Aircrew Automated Escape Systems," dated 15 May 1976.

To get an overview of NAVAIR resource availability, consider Figure 2-1. This figure delineates resource availability into four major groups: (1) System Availability, (2) Ejection Rationale, (3) Search and Rescue History, and (4) Medical Profile. Each of these, as the figure shows, is further subdivided into a series of binary decision branches. Stability of this resource model is realized whenever resource input, here interpreted as new crewmen, new carriers, new aircraft and new SAR equipment and personnel, equals resource output, interpreted here as obsolescence, outmoded techniques, injured crewmen, depreciation of equipment, and attrition of personnel.

To assist in developing an in-depth understanding of the medical profile portion of the entire system, a statistical analysis flow diagram, depicted in Figure 2-2, was derived. Not all statistical techniques shown in that figure were used in this report. A streamlined version of the techniques used in this report is illustrated in Figure 2-3.

The first technique applied to the data was the Two-Way Analysis of Variance, one observation per cell. This statistical analysis technique was applied specifically to fatalities incurred in connection with aircrew ejection from U.S. Navy high performance aircraft. Results are detailed in the next section.

Another useful technique applied to the data is that of non-parametric tests. The two tests employed here are the trend test and the run test. The trend test was applied to all A-4, A-5, A-6, A-7, F-4, F-8, and F-9 injury data. The run test was applied to all ejection related fatality data from

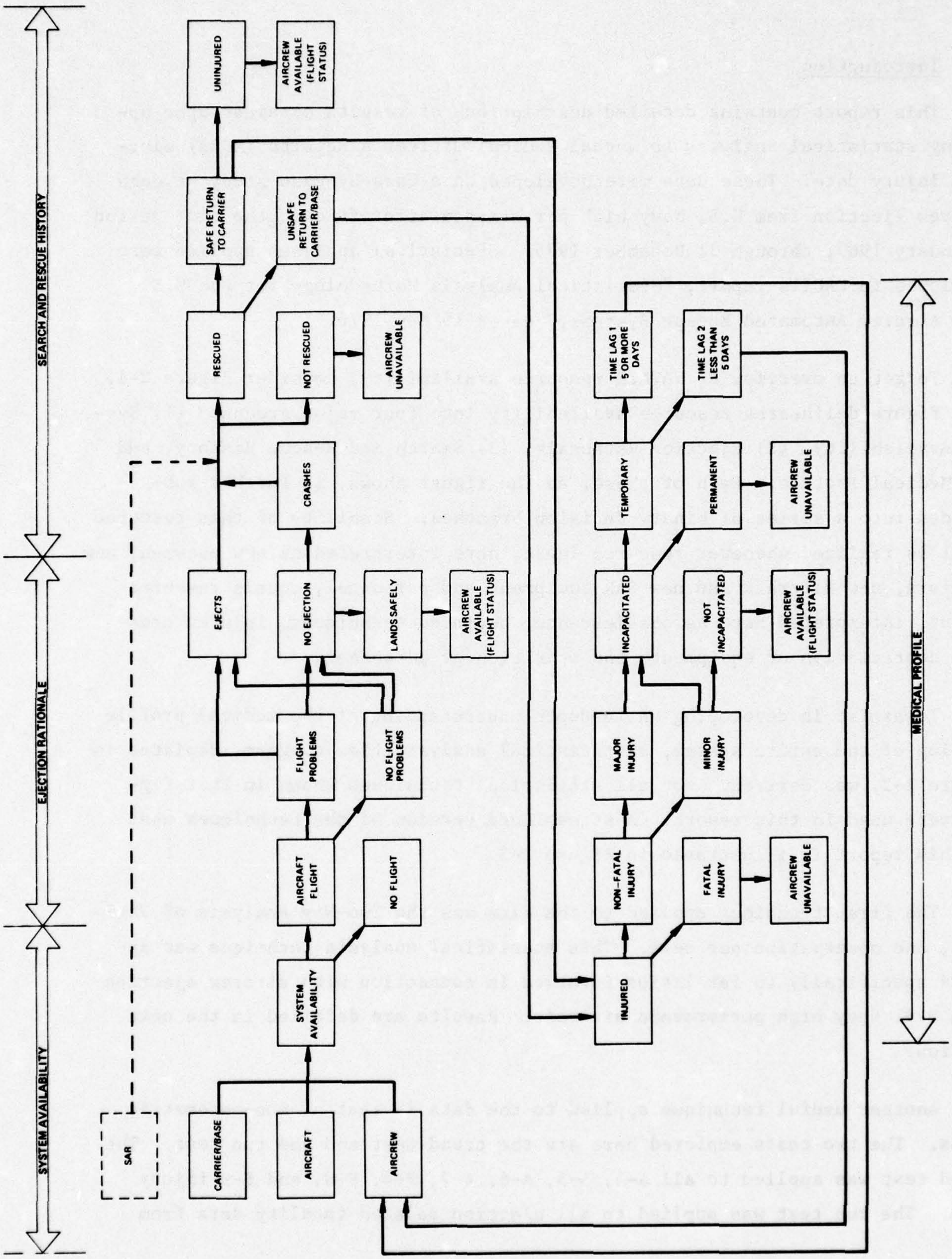

FIGURE 2-1. NAVAIR RESOURCE AVAILABILITY FOR HIGH PERFORMANCE AIRCRAFT

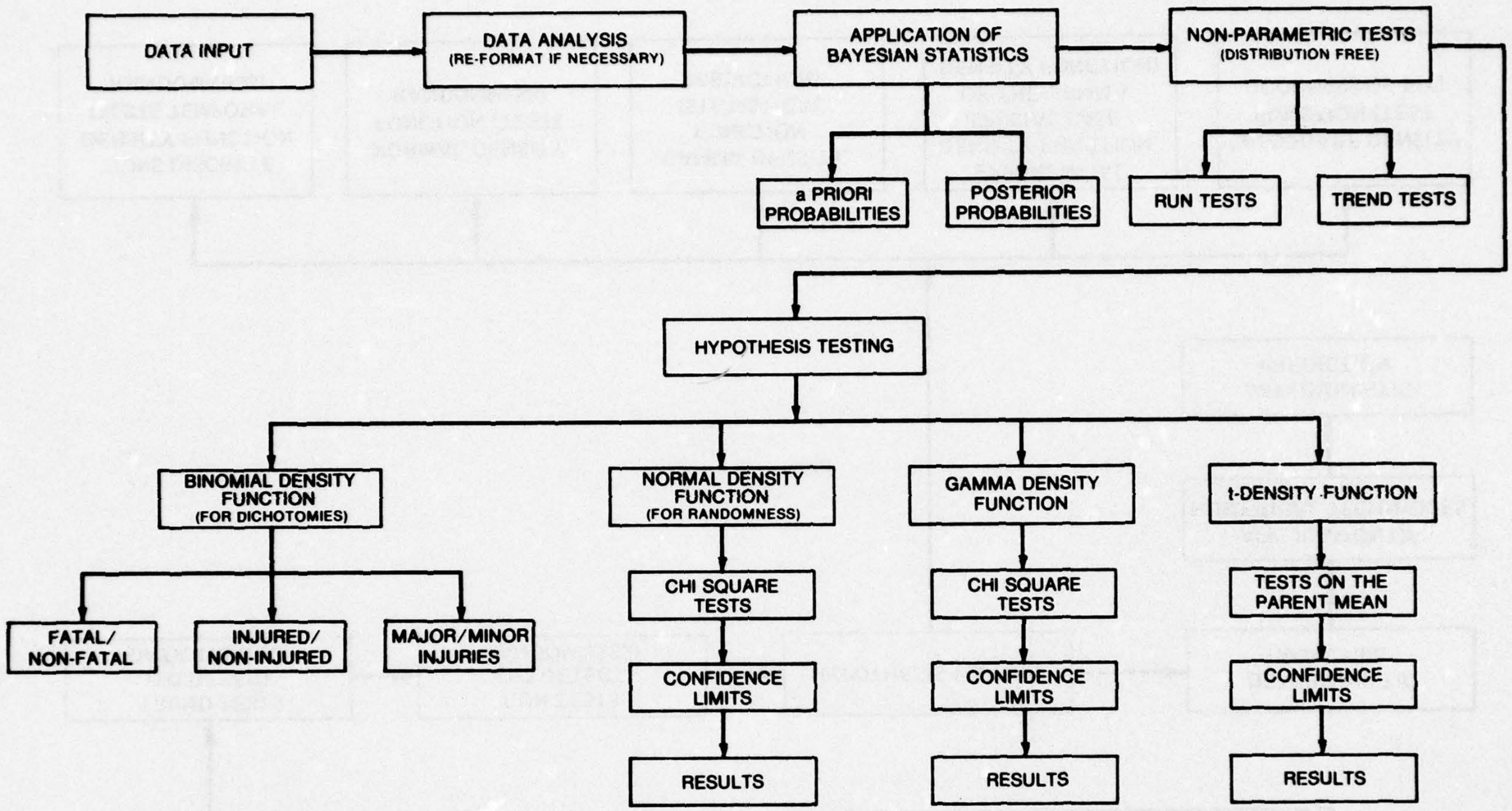

FIGURE 2-2. STATISTICAL ANALYSIS OF A SINGLE ATTRIBUTE DATA STREAM

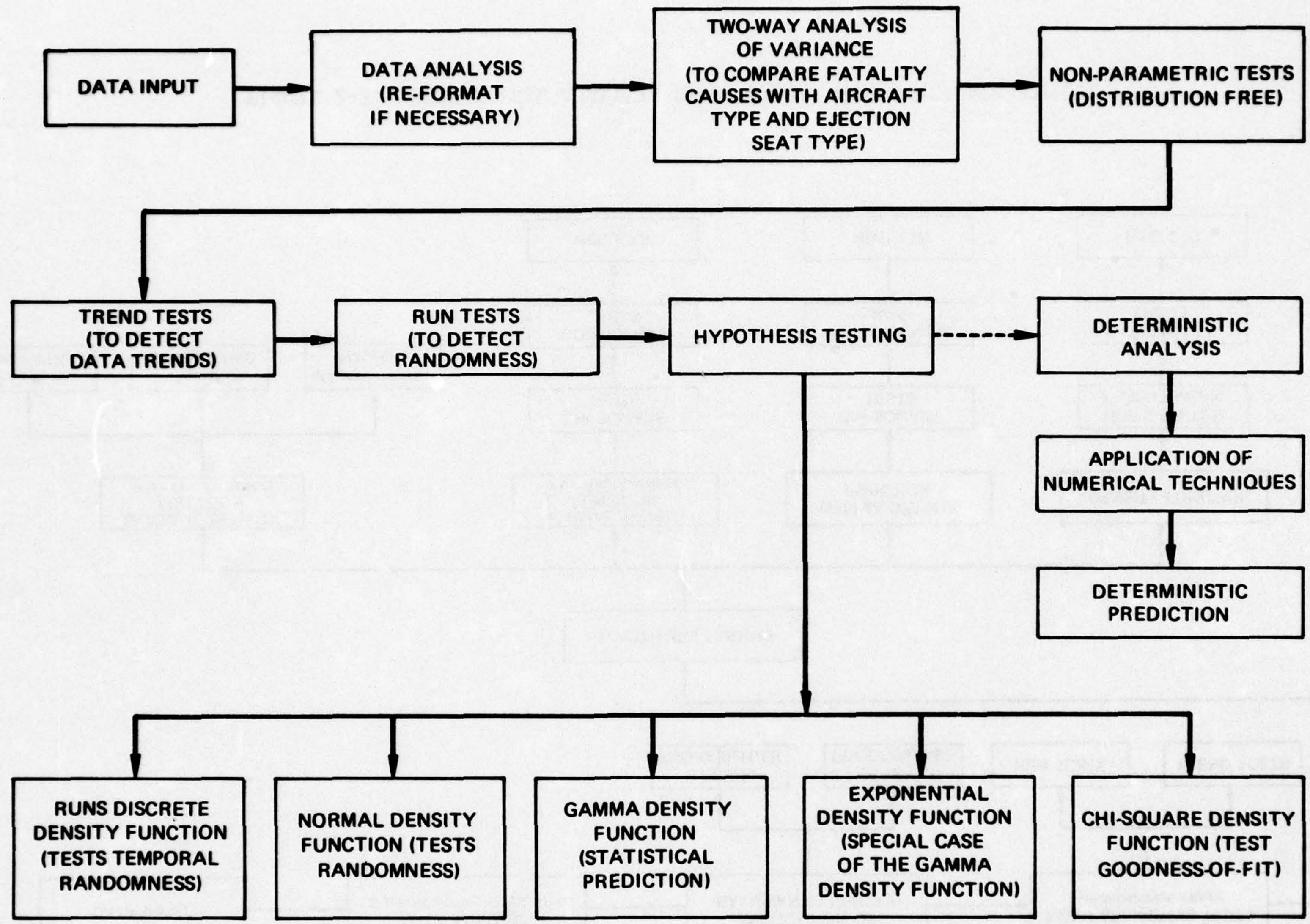

FIGURE 2-3. STATISTICAL AND DETERMINISTIC ANALYSES FOR SINGLE ATTRIBUTE DATA STREAMS

A-4, A-6, A-7, F-4, and F-8 aircraft. Results from the run test strongly suggest that A-6 fatalities do not occur randomly. Accordingly, a separate section is devoted exclusively to an analysis of A-6 fatalities.

One desirable end product of analysis is the ability to predict future patterns. Although some hypotheses were formulated and tested in the analyses already mentioned, it was believed appropriate to formulate and test additional hypotheses. These relate to determining the parent probability density functions from which the various ejection injury samples were extracted. If the parent density function can be derived, future injury patterns can be predicted. Three such ejection parent probability density functions were derived: (1) for the A-4 injuries, (2) the A-6 injuries, and (3) all ejection injuries from seven high performance aircraft taken collectively. All three parent density functions were gamma probability density functions. The A-4 injury pattern and the collective aircraft injury pattern were samples extracted from the exponential form of the gamma density function. Goodness-of-fit was tested with the Chi-square statistic.

This report is concluded by a summary of all results obtained to date.

Section 3.0

TWO-WAY ANALYSIS OF VARIANCE APPLIED TO FATALITIES EXPERIENCED UPON EJECTION FROM SELECTED U.S. NAVY HIGH PERFORMANCE AIRCRAFT

3.0 Two-Way Analysis of Variance Applied to Fatalities Experienced Upon Ejection from Selected U.S. Navy High Performance Aircraft

Analysis of variance is a statistical technique which can be used to make inferences about a set of observations, previously classified according to two criteria, displayed in a rectangular array. For example, it can be used to analyze a rectangular array consisting of averages of ejection related fatalities caused by (1) hardware, (2) aircrew, and (3) environment versus the ejections which occur from the following aircraft: (1) A-4, (2) A-7, (3) A-6, (4) F-4, and (5) the F-8. Examples of the three primary causes of fatalities are the following:

- Hardware: (a) Struck by personal equipment
  (b) Parachute dragging
- Aircrew: (a) Attempted ejection outside of safety envelope
  (b) Misuse of survival equipment
- Environment: (a) Windblast
  (b) Initial impact with the terrain.

A separate Two-Way Analysis of Variance was performed when the following ejection systems were considered: (1) ESCAPAC and (2) Martin-Baker. The three primary causes of fatalities (hardware, aircrew, and environment) remain unchanged. Each of the above analyses will be discussed separately.

Raw input data to the analyses are shown in Table 3-1 taken from the Medical Officer's Reports (MORs). Data from Table 3-1 were categorized and are summarized in Table 3-2. These were raw input to a Two-Way Analysis of Variance routine, one observation per cell. Data were then normalized to 100 ejections as shown in Table 3-3, and these numbers were the refined input to the Analysis of Variance routine.

Equations used in the computations are summarized below:

$$r = \text{number of rows in Table 3-3}$$

$$c = \text{number of columns in Table 3-3}$$

TABLE 3-1. FATALITY/LOST HISTORY UPON EJECTION FROM U.S. NAVY HIGH PERFORMANCE AIRCRAFT

	CAUSE OF FATALITY*	ESCAPAC			MARTIN-BAKER SEAT							TOTAL M-B AND ESCAPAC
		A-4	A-7	TOTAL	A-6	F-4	F-8	F-9	T-1	AV-8A	TOTAL	
PRE-EJECTION												
Lack of /Failure/ Misuse of Survival Equipment	A	0	0	0	0	0	1	0	0	0	1	1
Impact Forces	H	5	2	7	1	4	1	0	0	0	6	13
EJECTION												
Struck by Personal Equipment	H	1	0	1	0	0	0	0	0	0	0	1
Windblast	E	1	0	1	0	1	0	0	0	0	1	2
Initial Impact With Terrain	E	9	1	10	3	3	1	0	0	0	7	17
Struck By Ejection Seat After Landing on Ground	H	0	1	1	0	0	0	0	0	0	0	1
Parachute Dragging	H	1	0	1	0	0	0	0	0	0	0	1
Entanglement With Shroud Lines In Water	H	0	1	1	1	1	0	0	0	0	2	3
Outside Escape System Envelope	A	9	6	15	6	13	3	4	2	1	29	44
Lack of /Failure/ Misuse of Survival Equipment	A	3	1	4	2	0	1	0	0	0	3	7
Poor Body Position	A	0	1	1	0	0	0	0	0	0	0	1
Struck External Surface of Aircraft	H	0	0	0	1	0	0	0	0	0	1	1
Ejection Force	E	0	0	0	0	1	0	0	0	0	1	1
TOTAL FATALITIES		29	13	42	14	23	7	4	2	1	51	93
TOTAL NUMBER LOST	A,E,H	8	5	13	7	20	4	1	0	0	32	45
TOTAL FATALITIES AND LOST		37	18	55	21	43	11	5	2	1	83	138
TOTAL EJECTIONS		241	165	406	107	291	102	35	6	3	544	950
LOST												
Lost Out of Envelope		4	3	7	4	4	3	1	0	0	12	19
Marginal Out of Envelope		0	0	0	1	6	0	0	0	0	7	7
Possibly Out of Envelope		0	0	0	0	2	0	0	0	0	2	2
Also Out of Envelope		2	1	3	5	2	0	0	0	0	7	10
Possibly Out of Envelope		3	1	4	0	0	0	0	0	0	0	4
TOTAL OUT OF ENVELOPE												44 + 42 = 86

*A = Aircrew; E = Environment; H = Hardware.

TABLE 3-2. FATALITY HISTORY UPON EJECTION FROM U.S. NAVY HIGH PERFORMANCE AIRCRAFT (Actual Data)

EJECTION SEAT → / OTHER CAUSES ↓	ESCAPAC EJECTION SEAT			MARTIN-BAKER EJECTION SEAT				GRAND TOTAL
AIRCRAFT CAUSE → / OTHER CAUSES ↓	A-4	A-7	TOTAL ESCAPAC FATALITIES	A-6	F-4	F-8	TOTAL MARTIN-BAKER FATALITIES	
Hardware	10	6	16	5	12	2	19	35
Aircrew	15	9	24	11	19	7	37	61
Environment	12	3	15	5	12	2	19	34
TOTAL:	37	18	55	21	43	11	75	130
EJECTIONS	241	163	404	107	291	102	500	904

TABLE 3-3. TWO-WAY ANALYSIS OF VARIANCE (One Observation per Cell) AIRCRAFT CAUSES VERSUS HARDWARE, AIRCREW, AND ENVIRONMENT CAUSES OF FATALITIES (Per 100 Ejections)

AIRCRAFT CAUSES → / OTHER CAUSES ↓	A-4	A-7	A-6	F-4	F-8	TOTAL
Hardware	4	4	5	4	2	19
Aircrew	6	6	10	7	7	36
Environment	5	2	5	4	2	18
TOTAL:	15	12	20	15	11	73

$$\left.\begin{aligned} R_i &= \sum_{j=1}^{c} x_{ij} \\ C_j &= \sum_{i=1}^{r} x_{ij} \\ G &= \sum_{i=1}^{r} \sum_{j=1}^{c} x_{ij} \end{aligned}\right\} \tag{3.1}$$

x_{ij} = number in the ij^{th} cell in the rectangular array (Table 3-3, for example).

$$SSR = \sum_{i=1}^{r} \frac{R_i^2}{c} - \frac{G^2}{rc} \tag{3.2}$$

$$SSC = \sum_{j=1}^{c} \frac{C_i^2}{r} - \frac{G^2}{rc} \tag{3.3}$$

$$SSE = \sum_{i=1}^{r} \sum_{j=1}^{c} x_{ij}^2 - \sum_{i=1}^{r} \frac{R_i^2}{c} - \sum_{j=1}^{c} \frac{C_j^2}{r} + \frac{G^2}{rc} \tag{3.4}$$

$$SST = \sum_{i=1}^{r} \sum_{j=1}^{c} x_{ij}^2 - \frac{G^2}{rc} \tag{3.5}$$

Numerical values from Table 3-3 were inserted into equations (3.1) - (3.5) and the output is summarized in Table 3-4.

The following null hypotheses were formulated using a common denominator of 100 ejections from each aircraft:

$H_{o,r}$: There is no difference in the average number fatalities caused by hardware hazards, aircrew actions, and environmental conditions.

$H_{o,c}$: There is no difference in the average number of fatalities caused by ejection from the A-4, A-7, A-6, F-4, and F-8 aircraft.

The null hypotheses for the rows, $H_{o,r}$, and the columns, $H_{o,c}$, were tested at the $\alpha = 0.05$ level of significance. Results follow in Table 3-4.

TABLE 3-4. TWO-WAY ANALYSIS OF VARIANCE--AIRCRAFT EJECTION FATALITIES CAUSED BY HARDWARE, AIRCREW, AND ENVIRONMENTAL CONDITIONS UPON EJECTION FROM VARIOUS U.S. NAVY HIGH PERFORMANCE AIRCRAFT

	Degrees of Freedom	Sum of Squares	Mean Square	F-Statistic (Computed)
Between Rows	(r-1)	(SSR)	(SSR)/(r-1)	MSR/MSE
	2	40.93	20.47	19.50
Between Columns	(c-1)	(SSC)	(SSC)/(c-1)	MSC/MSE
	4	16.40	4.1	3.90
Residual	(r-1)(c-1)	(SSE)	$\frac{(SSE)}{(r-1)(c-1)}$	
	8	8.4	1.05	
TOTAL:	(rc-1)	(SST)		
	14	65.73		

	Level of Significance	Critical F-Statistic	Computed F-Statistic	Inference
Rows	$\alpha = 0.05$	$F_{o,r}(2,8) = 4.46$	$\hat{F}_{o,r}(2,8) = 19.50$	Reject $H_{o,r}$
Columns	$\alpha = 0.05$	$F_{o,c}(4,8) = 3.84$	$\hat{F}_{o,c}(4,8) = 3.90$	Reject $H_{o,c}$

From these results, the conclusion is reached that there is a significant difference in the average number of fatalities caused by hardware hazards, aircrew actions, and environmental conditions. Moreover, a significant difference exists in the average number of fatalities which occur upon ejection from the various aircraft studied.

A completely analogous analysis was performed for comparison of fatalities caused by hardware hazards, aircrew actions, and environmental conditions whenever ejections were initiated from two ejections seat systems: (1) ESCAPAC, and (2) the Martin-Baker ejection seat. Refined input to the analysis of variance procedure is displayed in Table 3-5. Numerical results, obtained from equations (3.1) - (3.5) are shown in Table 3-6.

The following null hypotheses were formulated:

$H_{o,r}$: There is no difference in the average number of fatalities caused by hardware hazards, aircrew actions, and environmental conditions.

$H_{o,c}$: There is no difference in the average number of fatalities caused upon ejection from either the ESCAPAC ejection seat or the Martin-Baker ejection seat.

The null hypotheses for the rows, $H_{o,r}$, and the columns, $H_{o,c}$, were tested at the $\alpha = 0.05$ level of significance. Results follow in Table 3-6.

TABLE 3-5. TWO-WAY ANALYSIS OF VARIANCE (One Observation Per Cell) EJECTION SEAT CAUSES VERSUS HARDWARE, AIRCREW, AND ENVIRONMENT CAUSES OF FATALITIES PER 500 EJECTIONS

EJECTION SEAT → / ↓ CAUSE OF FATALITY	ESCAPAC	MARTIN-BAKER	TOTAL
Hardware	20	19	39
Aircrew	30	37	67
Environment	19	19	38
TOTAL:	69	75	144

TABLE 3-6. TWO-WAY ANALYSIS OF VARIANCE--EJECTION SEAT FATALITIES CAUSED BY HARDWARE, AIRCREW, AND ENVIRONMENTAL CONDITIONS UPON EJECTION, USING VARIOUS EJECTION SEATS, IN U.S. NAVY HIGH PERFORMANCE AIRCRAFT PER 500 EJECTIONS

	Degrees of Freedom	Sum of Squares	Mean Square	F-Statistic (Computed)
Between Rows	(r-1)	(SSR)	(SSR)/(r-1)	MSR/MSE
	2	271	135.5	14.26
Between Columns	(c-1)	(SSC)	(SSC)/(c-1)	MSC/MSE
	1	6	6	0.63
Residual	(r-1)(c-1)	(SSE)	$\frac{(SSE)}{(r-1)(c-1)}$	
	2	19	9.5	
TOTAL	(rc-1) 5	(SST) 296		

	Level of Significance	Critical F-Statistic	Computed F-Statistic	Inference
Rows	$\alpha = 0.05$	$F_{o,r}(2,2) = 19.0$	$\hat{F}_{o,r}(2,2) = 14.26$	Accept $H_{o,r}$
Columns	$\alpha = 0.05$	$F_{o,c}(1,2) = 18.5$	$\hat{F}_{o,c}(1,2) = 0.63$	Accept $H_{o,c}$

From these results, the conclusion is reached that there is no significant difference in the average number of fatalities caused by hardware hazards, aircrew actions, and environmental conditions. No difference was discovered when a comparison was made of the two ejection system types.

An apparent discrepancy exists when comparing the results of the above Analysis of Variance tests. In one case a significant difference exists, and

in another practically analogous case, there is no significant difference in injury patterns. Further investigation is required to account for this phenomenon. It may be that the data tabulation, the data classification, or the data interpretation is in error. This might explain the inconsistency. In any event, under the assumptions cited at the beginning of this section, there seems to be, in certain cases, some dependency among the variables analyzed.

Section 4.0

NON-PARAMETRIC STATISTICAL TESTS

4.0 Non-Parametric Statistical Tests

Two non-parametric statistical tests of particular importance in the present investigation are: (1) the one sample trend test, and (2) the one sample run test. The trend test was applied to all A-4, A-5, A-6, A-7, F-4, F-8, and F-9 ejection injury data acquired over the time period 1 January 1969, through 31 December 1975. Non-parametric run tests were applied to ejection related fatality data, including "lost and unknown," for the A-4, A-6, A-7, F-4 and F-8 aircraft. It was discovered that on a chronological time basis, fatalities do not occur randomly upon ejection from A-6 aircraft. Consequently, a separate section in this report has been devoted to an analysis of A-6 ejection related fatalities.

4.1 Non-Parametric One Sample Trend Analysis of A-4 Ejection Related Injury Data

A trend analysis is a non-parametric statistical test used to discover whether the data sample under investigation contains an underlying trend. The procedure is straightforward: Select the first data point in the sample and compare its magnitude with the magnitude all successive data points. Count and record the number of times its magnitude exceeds the magnitude of all other data points. Continue in this manner counting and recording the number of times $x_i > x_j$ for $i < j$. Tabulate the results, then add the number of "trends" so obtained. Finally, apply normal density function theory to infer a conclusion about an underlying trend in the data sample.

To apply this to a specific example, Table 4-1 was constructed from actual MORs data. To quantify MORs data, the following correspondences were established:

MORs Data	Data in Table 4-1
G	1
F	2
B	3
A	4
L & U	5

TABLE 4-1. NON-PARAMETRIC ONE SAMPLE TREND ANALYSIS OF A-4 EJECTION RELATED INJURY DATA

2, 1, 2, 1, 1, 1, 1, 2, 3, 5, 1, 1, 1, 4, 2, 2, 2, 2, 2, 2, 2, 1, 2,
2, 2, 3, 1, 2, 1, 4, 1, 4, 1, 2, 1, 3, 2, 3, 4, 2, 1, 1, 2, 2, 4, 3,
1, 1, 1, 4, 1, 2, 1, 1, 3, 1, 2, 1, 1, 1, 2, 4, 2, 1, 3, 3, 4, 1, 2,
2, 4, 2, 2, 1, 5, 1, 2, 2, 1, 2, 3, 4, 2, 3, 1, 2, 1, 3, 1, 2, 2, 3,
2, 1, 1, 4, 1, 2, 2, 2, 1, 2, 1, 3, 1, 1, 1, 1, 3, 1, 3, 1, 1, 3, 3,
1, 1, 1, 3, 2, 2, 2, 1, 3, 3, 1, 2, 3, 1, 1, 4, 1, 2, 4, 1, 1, 2, 3,
4, 5, 1, 1, 1, 1, 2, 3, 4, 2, 1, 4, 3, 2, 2, 3, 3, 1, 1, 1, 1, 1, 3,
2, 1, 2, 1, 3, 3, 1, 1, 4, 4, 4, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 2, 1, 1, 3, 1, 5, 1, 2, 1, 1, 1, 3, 1, 4, 4, 2, 5, 1, 1, 1, 4, 1,
1, 1, 2, 3, 4, 1, 4, 1, 2, 4, 1, 5, 1, 4, 3, 1, 4, 1, 3, 1, 3, 1, 1,
2, 2, 5, 1, 2, 5, 1, 1, 2, 4, 4.

Now formulate the null hypothesis:

H_o: The A-4 Ejection Injury data does not contain an underlying trend.

To test this hypothesis at the $\alpha = 0.05$ level of significance, first compute N_1 and N_2, where

$$\left.\begin{aligned} N_1 &= -\,1.96\,\sigma_y + \mu_y \\ N_2 &= +\,1.96\,\sigma_y + \mu_y \end{aligned}\right\} \qquad (4.1)$$

and

$$\left.\begin{aligned} \mu_y &= \frac{N(N-1)}{4} \\ \sigma_y^2 &= \frac{2(N)^3 + 3(N)^2 - 5(N)}{72} \end{aligned}\right\} \qquad (4.2)$$

N = number of data points under analysis, in this case $N = 241$.

From equations (4.1) and (4.2), $N_1 = 13{,}590.5$, $N_2 = 15{,}329.5$, and from Table 4-2, $N_c = 10{,}066$. Here, N_c is the number of reverse arrangements in the data sample. The acceptance region is $N_1 < N_c < N_2$. Since $N_c < N_1$, then clearly the null hypothesis is rejected, and we conclude that with 95 percent confidence, the A-4 ejection related injury data does contain an underlying trend.

A flow diagram illustrating procedure for implementing a non-parametric trend analysis on a digital computer is displayed in Figure 4-1.

TABLE 4-2. TRENDS OBSERVED IN A NON-PARAMETRIC ONE SAMPLE TREND ANALYSIS OF A-4 EJECTION RELATED INJURY DATA

109, 0, 108, 0, 0, 0, 0, 104, 160, 224, 0, 0, 0, 193, 101, 101, 101,
101, 101, 101, 101, 0, 100, 100, 100, 147, 0, 99, 0, 174, 173, 0, 96,
0, 140, 95, 139, 170, 95, 0, 0, 93, 93, 165, 134, 0, 0, 0, 161, 0,
89, 0, 0, 127, 0, 86, 0, 0, 0, 83, 150, 83, 0, 119, 119, 146, 0, 81,
81, 143, 81, 81, 0, 160, 0, 79, 79, 0, 78, 108, 134, 78, 107, 0, 77,
0, 104, 0, 75, 75, 101, 75, 0, 0, 123, 0, 72, 72, 72, 0, 71, 0, 91,
0, 0, 0, 0, 87, 0, 86, 0, 0, 84, 84, 0, 0, 0, 81, 60, 60, 0, 77, 77,
0, 58, 75, 0, 0, 89, 0, 55, 87, 0, 0, 53, 68, 83, 83, 0, 0, 0, 0, 49,
63, 77, 49, 0, 75, 61, 48, 48, 59, 59, 0, 0, 0, 0, 0, 54, 43, 0, 42,
0, 50, 50, 0, 0, 56, 56, 56, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
26, 0, 0, 32, 0, 48, 0, 22, 0, 0, 0, 26, 0, 29, 29, 18, 36, 0, 0, 0,
25, 0, 0, 0, 12, 17, 20, 0, 19, 0, 10, 17, 0, 20, 0, 15, 12, 0, 13,
0, 10, 0, 9, 0, 0, 3, 3, 7, 0, 2, 5, 0, 0, 0, 0, 0.

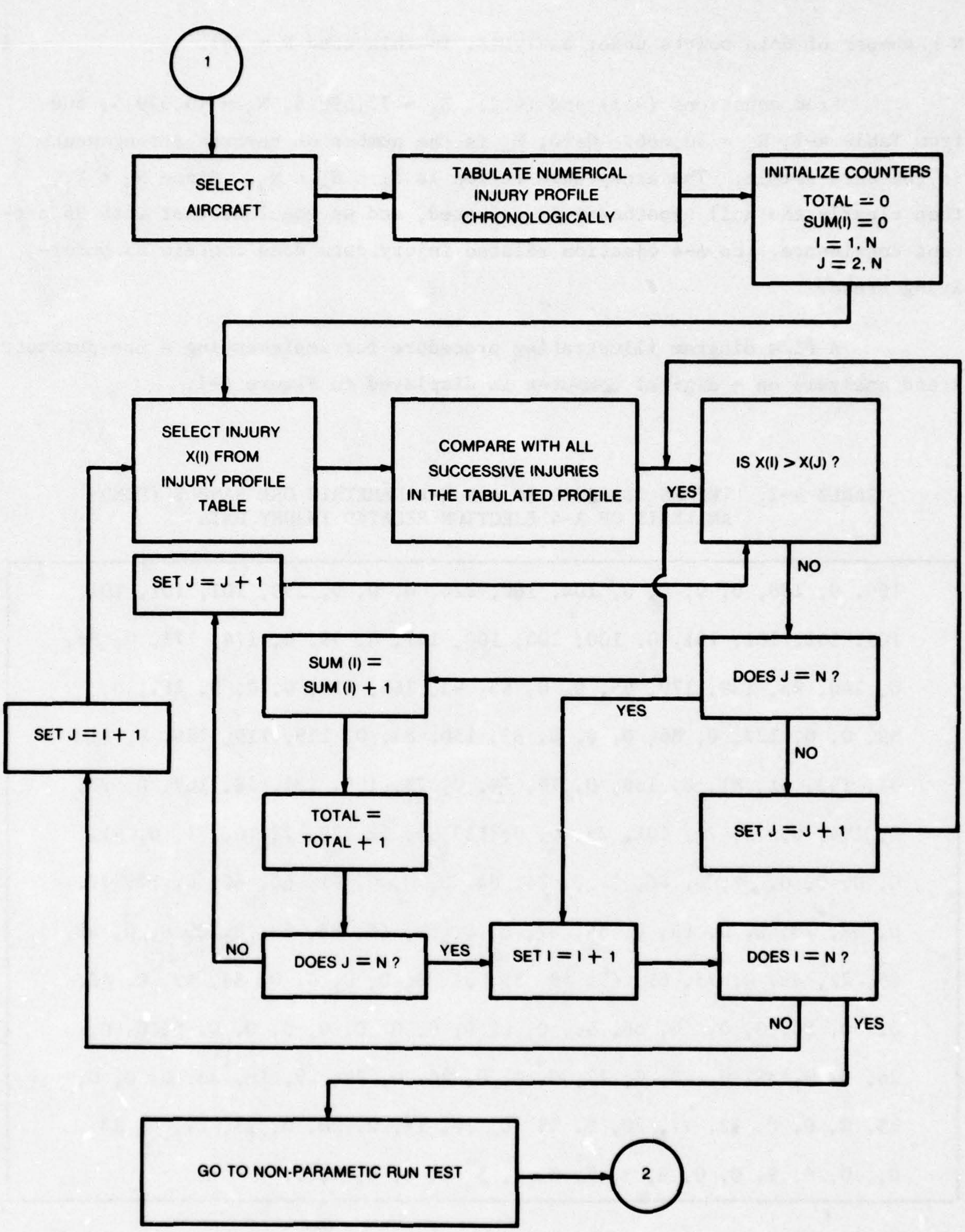

FIGURE 4-1. NON-PARAMETRIC STATISTICAL TREND TEST

4.2 Non-Parametric One Sample Run Test Applied to A-4 Ejection Related Fatality Data

One statistical technique that can be used to test randomness in a chronologically stored data stream is that of the one-sample run test. A dichotomous situation needs to be constructed prior to application of the test. A run is defined as a set of symbols of one kind preceded and followed by symbols of another kind or no symbols at all.

For the high performance aircraft cited in the MORs, there is a minimum of six dichotomies which can be constructed from data from each aircraft listed. The aircraft selected for this analysis was the A-4, and the dichotomy of interest is that of fatalities versus all other injuries. As mentioned in Table 4-3, a 1 represents all injuries except known fatalities, and a 0 represents known fatalities. A lost or unknown ejection also was set to a 1.

TABLE 4-3. AN ANALYSIS OF A RUN TEST FOR A-4 DICHOTOMOUS EJECTION DATA (1 = Non-Fatal Injury; 0 = Fatal Injury)

1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 0, 1, 0, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 1,
1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 0, 1, 1,
1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0, 1, 1, 1, 1,
0, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 1, 0, 0, 1, 1, 1, 1, 1, 0, 1,
1, 1, 1, 1, 0, 1, 0, 1, 1, 0, 1, 1, 1, 0, 1, 1, 0, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0.

The null hypothesis to be tested is the following:

H_o: Upon ejection from A-4 aircraft, known fatalities occur randomly.

The procedure for testing validity, or not, of H_o is straightforward. Count the number, n_1, of 1's, the number n_2 of 0's, and the number of runs, u. From Table 4-3, the following is observed:

$$\left. \begin{aligned} n_1 &= 212 \\ n_2 &= 29 \\ u &= 50 \end{aligned} \right\} \tag{4.3}$$

From the statistical theory of runs,

$$E(u) = \frac{2\, n_1\, n_2}{n_1 + n_2} + 1, \tag{4.4}$$

$$\mathrm{Var}\ (u) = \frac{2\, n_1\, n_2\ (2\, n_1\, n_2 - n_1 - n_2)}{(n_1 + n_2)^2\ (n_1 + n_2 - 1)}\ , \tag{4.5}$$

and if $u > 10$,

$$z_c = \frac{u - E(u)}{\sqrt{\mathrm{Var}\ (u)}}\ . \tag{4.6}$$

Substitute the numbers from equation (4.3) into equations (4.4) - (4.6) to get:

$$\left. \begin{aligned} E(u) &= 52.020 \\ \mathrm{Var}(u) &= 10.634 \\ z_c &= -\ 0.6194 \end{aligned} \right\} \tag{4.7}$$

From tabulated values,

$$\left. \begin{aligned} z_{-0.025} &= -\ 1.96 \\ z_{0.025} &= +\ 1.96. \end{aligned} \right\} \tag{4.8}$$

Since $z_{-0.025} < z_c < z_{+0.025}$, that is, since $-1.96 < -0.6194 < +1.96$, the null hypothesis of randomness cannot be rejected at the $\alpha = 0.05$ level of significance. That is, it can be said with 95 percent confidence that, from results of the run test, fatalities upon ejection from A-4 U.S. Navy high performance aircraft occur randomly. An overview computer flow diagram for computing run tests is shown in Figure 4-2.

This non-parametric statistical run test was applied to ejection related fatality data on the A-6, A-7, F-4, and F-8 aircraft. The major conclusion inferred from these analyses is that on a chronological time basis, ejection related fatalities associated with all aircraft mentioned above, except the A-6, occur randomly. A special analysis of A-6 fatalities is presented in Section 5.

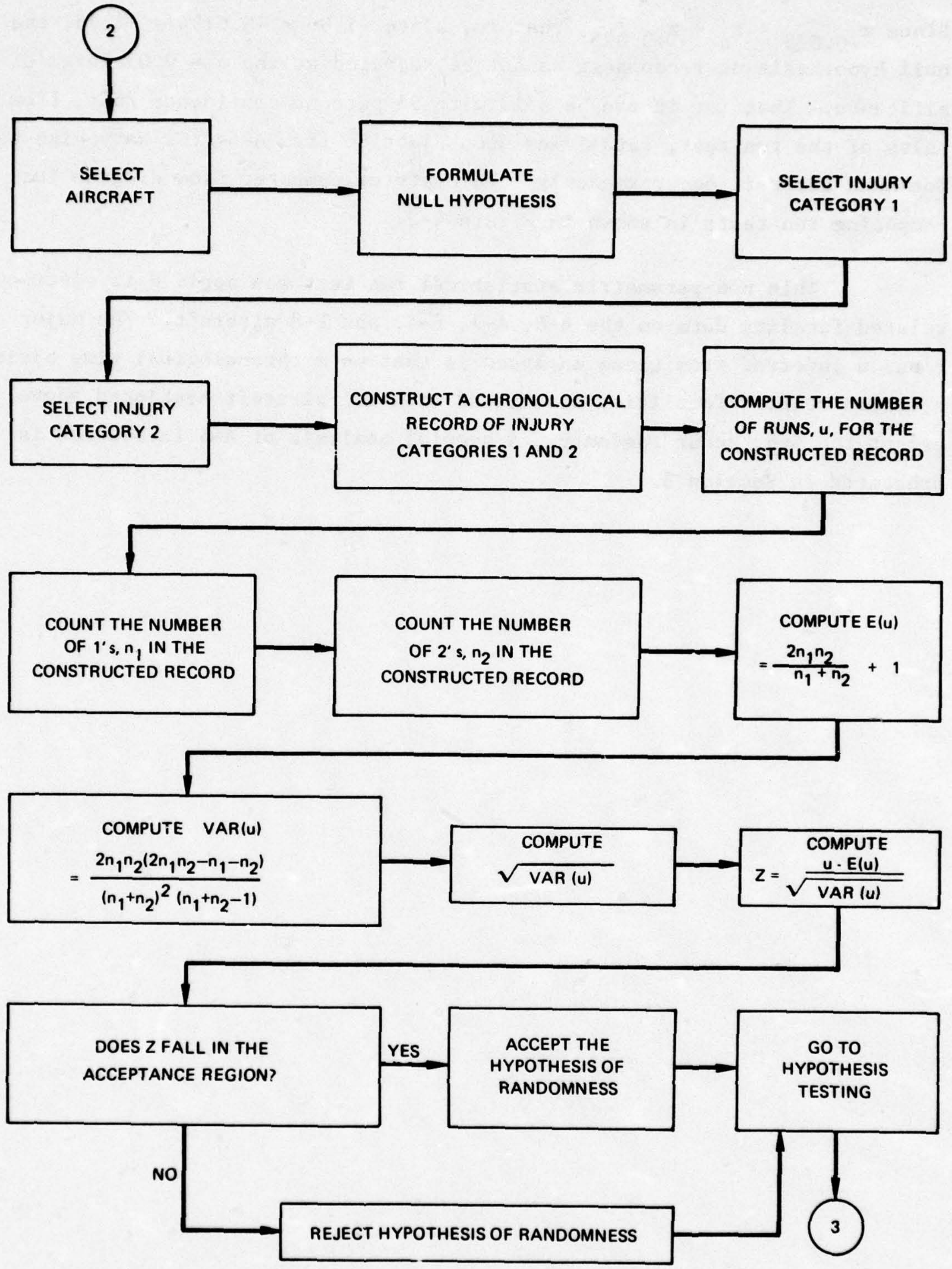

FIGURE 4-2. NON-PARAMETRIC STATISTICAL RUN TEST

Section 5.0

AN ANALYSIS OF A-6 EJECTION RELATED FATALITIES

5.0 An Analysis of A-6 Ejection Related Fatalities

A statistical test called the run test was applied to A-6 ejection fatality data so that an inference could be drawn from the following null hypothesis:

> H_o: Upon ejection from A-6 aircraft, known fatalities occur randomly on a chronological time basis.

Under the run test, this hypothesis was rejected at the $\alpha = 0.05$ level of significance. Stated another way, it can be said with 95 percent confidence that fatalities do not occur randomly upon ejection from the A-6 aircraft. An analysis to ascertain whether A-6 ejection fatalities occur in a deterministic pattern then was conducted. Predictions about the occurrence of future fatalities was then made from the given pattern. Here it must be emphasized that any prediction about future fatality patterns is predicated on the assumption that future ejection conditions are similar to present and past conditions.

5.1 A Deterministic Analysis of A-6 Ejection Related Fatality Patterns

To start the analysis, a tabulation of some information about A-6 fatality occurrences was constructed. These data are shown in Table 5-1. A graph of the total fatalities, T_i, versus decimal approximation of the date, t_i, clearly demonstrates that the data fall quite naturally into three linear clusters: L_1 from 7 March 1969 through 12 June 1970; L_2 from 27 May 1972 through 19 September 1973; and L_3 from 9 October 1974 through 20 August 1975.

An analysis was performed to fit each line segment to the given data points such that the fit was best in the least squares sense. The following equations were used to compute the slope and intercept, respectively, for each line segment.

$$m_j = \frac{\sum_{i=1}^{n_j} T_i t_i - n_j \bar{T}_j \bar{t}_j}{\sum_{i=1}^{n_j} t_i^2 - n_j \bar{t}_j^2} ; \left\{ \begin{array}{l} i = 1, 2, \ldots, n_j \\ \\ j = 1, 2, 3. \end{array} \right. \qquad (5.1)$$

$$b_j = \bar{T}_j - m_j \bar{t}_j \qquad (5.2)$$

TABLE 5-1. AN ANALYSIS OF A-6 EJECTION RELATED FATALITY DATA

Date	Day	Kind of Aircraft	Kind of Fatality	t_i = Date (Decimal Approximation) (0.00 = 1-1-69)	T_i = Total Fatalities
03-07-69	66	A-6A	A	0.1808	1.
03-07-69	66	A-6A	A	0.1808	2.
08-26-69	238	A-6A	A	0.6521	3.
12-26-69	360	A-6A	A	0.9863	4.
12-26-69	360	A-6A	L	0.9863	5.
02-16-70	47	A-6A	L	1.1288	6.
04-22-70	112	A-6A	A	1.3068	7.
06-12-70	163	A-6A	A	1.4466	8.
05-27-72	148	A-6E	A	3.4044	9.
08-09-72	222	KA-6D	A	3.6066	10.
10-24-72	298	A-6A	L	3.8142	11.
10-31-72	305	KA-6D	A	3.8334	12.
10-31-72	305	KA-6D	A	3.8334	13.
09-19-73	262	A-6A	A	4.7178	14.
09-19-73	262	A-6A	A	4.7178	15.
10-09-74	282	KA-6D	L	5.7726	16.
11-20-74	324	A-6A	A	5.8877	17.
11-20-74	324	A-6A	A	5.8877	18.
01-13-75	13	EA-6B	L	6.0356	19.
06-25-75	176	A-6E	L	6.4822	20.
08-20-75	232	KA-6D	L	6.6356	21.

Here,

$$\bar{T}_j = \sum_{i=1}^{n_j} T_i/n_j \quad \text{and} \quad \bar{t}_j = \sum_{i=1}^{n_j} t_i/n_j . \tag{5.3}$$

Apply the above to data in Table 5-1 to get:

$$L_1: \quad T_{L_1} = 4.93346\, t_{L_1} + 0.26413 , \tag{5.4}$$

$$L_2: \quad T_{L_2} = 3.79448\, t_{L_2} - 3.13869 , \tag{5.5}$$

$$L_3: \quad T_{L_3} = 4.93626\, t_{L_3} - 11.69458 . \tag{5.6}$$

Next, the centroids of the line segments were computed, graphed, and a curve drawn through them. Coordinates of the centroids are as follows:

$$\left.\begin{aligned} \bar{T}_{L_1} &= 4.5; \; \bar{t}_{L_1} = 0.8586 \\ \bar{T}_{L_2} &= 12.0; \; \bar{t}_{L_2} = 3.98966 \\ \bar{T}_{L_3} &= 18.5; \; \bar{t}_{L_3} = 6.1169 \end{aligned}\right\} \tag{5.7}$$

A graph of these points appeared to follow a quadratic of the form

$$\bar{T} = \sum_{k=0}^{2} A_k \bar{t}^k . \tag{5.8}$$

By using an arithmetic averaging scheme, it was easy to compute A_k in equation (5.8). The complete quadratic is:

$$\bar{T} = 2.8734476 + 1.786602\, \bar{t} + 0.125562\, \bar{t}^2 \tag{5.9}$$

To project into the future, it is desired to find the centroid of another line segment L_4. To accomplish this, notice that the time distance between the derived centroids of line segments L_1, L_2, and L_3 appears to contract according to a logarithmic law. Thus,

$$\Delta\bar{t} = B_1 e^{B_2 \bar{\bar{t}}}, \tag{5.10}$$

where

$$\Delta\bar{t} = \bar{t}_{j+1} - \bar{t}_j \quad \text{and} \quad \bar{\bar{t}} = \frac{\bar{t}_{j+1} + \bar{t}_j}{2},$$

and B_1, B_2 are constants to be determined. Data in Table 5-2 will assist in computing B_1 and B_2.

TABLE 5-2. DATA USED FOR COMPUTING CONSTANTS B_1 AND B_2 IN AN EXPONENTIAL EQUATION

$\bar{t} = (t_{j+1} + t_j)/2$	$\Delta\bar{t} = \bar{t}_{j+1} - \bar{t}_j$	$\bar{\bar{t}} = (\bar{t}_{j+1} + \bar{t}_j)/2$
0.8586		
	3.13106	2.42413
3.98966		
	2.12724	5.05328
6.1169		

The data in Table 5-2 permit equation (5.10) to be written thus:

$$\Delta\bar{t} = 4.471717\, e^{-0.147023\, \bar{\bar{t}}} \tag{5.11}$$

From the definitions given in the headings of Table 5-2, it is easy to derive the following iterative scheme:

$$D_1 = \bar{\bar{t}} - 6.1169 - 2.23586\, e^{-0.147023\, \bar{\bar{t}}} \tag{5.12}$$

$D \doteq 0$ whenever $\bar{\bar{t}} = 6.92475$. Thus,

$$\bar{t}_2 = (2)\ (\bar{\bar{t}}) - \bar{t}_1 ,$$

or

$$\bar{t}_2 = (2)(6.92475) - 6.1169$$

$$\bar{t}_2 = 7.7326$$

From equation (5.9),

$$\bar{T}\ (7.7326) = 24.1964 \tag{5.13}$$

Thus,

$$\left.\begin{aligned} \bar{t}_2 &= 7.7326 \\ \bar{T} &= 24.1964 \end{aligned}\right\} \tag{5.14}$$

These are the extrapolated coordinates of the projected centroid for line segment L_4. The slope of L_4 is assumed to be the mean of the slopes of lines L_1, L_2, and L_3. Thus,

$$m_4 = \sum_{j=1}^{3} m_j/3 \tag{5.15}$$

$$m_4 = (4.93346 + 3.79448 + 4.93626)/3$$

$$m_4 = 4.554734 \tag{5.16}$$

From equations (5.14) and (5.16), the line segment L_4 is completely determined. Thus,

$$L_4\!: \ T_{L_4} = 4.554734\ t - 11.023536\ . \tag{5.17}$$

To determine a reasonable starting point on the line L_4, again assume that the time lengths between the last points on preceding line segments and first points on current line segments is logarithmic in nature. That is,

$$\Delta t = c_1 \, e^{c_2 \bar{t}} \quad . \tag{5.18}$$

Proceed in a manner entirely analogous to that developed earlier, hence write an iterative scheme as follows:

$$D_2 = \bar{t} - 6.6356 - 1.6664237 \, e^{-0.2193389 \, \bar{t}} \tag{5.19}$$

$D_2 \doteq 0$ whenever $\bar{t} = 6.9949 = (t_1 + t_2)/2$, where $t_1 = 6.6356$. Thus, $t_2 = 7.3542$. Translated, this becomes the 7^{th} year, which is 1976, and (0.3542) (366) = 130 days into 1976. That is, 15 May 1976, which is a predicted start date for fatalities clustered about the line segment L_4. Average time length per line segment = 1.1474 years. Time length of line segment L_4 = 1.1458 years over the time interval 15 May 1976 to 1 July 1977 inclusive.

A prediction scheme, subject to the assumptions stated earlier, is hypothesized in Table 5-3 below. These dates should be interpreted as "on or about" dates since bounds on the dates given for line segments L_1, L_2, and L_3 have not yet been established. Moreover, the dates given fall on the line segment L_4. A graphical display of these results is shown in Figure 5-1.

TABLE 5-3. PREDICTED DATES FOR A-6 EJECTION RELATED FATALITIES, UNDER THE ASSUMPTION OF LOGARITHMIC TIME CONTRACTION

Predicted		Observed	
Date	Fatalities	Date	Fatalities
May 15, 1976	2		
August 15, 1976	1		
December 1, 1976	1		
February 15, 1977	1		
May 1, 1977	1		
July 25, 1977	1		

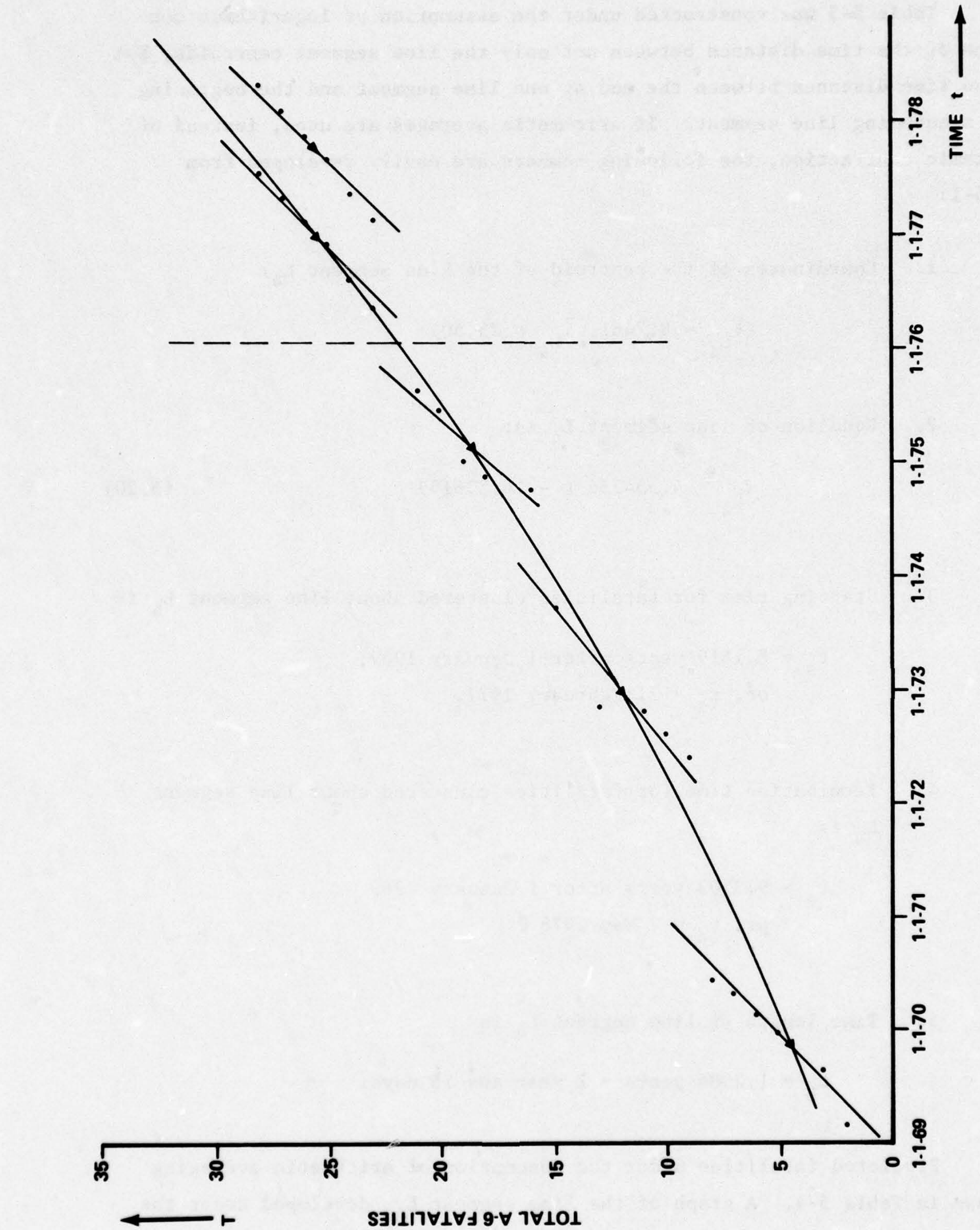

FIGURE 5-1. GRAPH OF TOTAL A-6 EJECTION RELATED FATALITIES VERSUS TIME

Table 5-3 was constructed under the assumption of logarithmic contraction of the time distance between not only the line segment centroids, but also the time distance between the end of one line segment and the beginning of the succeeding line segment. If arithmetic averages are used, instead of logarithmic contraction, the following numbers are easily developed from Table 5-1:

1. Coordinates of the centroid of the Line Segment L_4:

$$(\bar{t}_{L_4} = 8.7461,\ \bar{T}_{L_4} = 25.50)$$

2. Equation of line segment L_4 is:

$$T_{L_4} = 4.554734\ t - 14.336159 \qquad (5.20)$$

3. Starting time for fatalities clustered about line segment L_4 is

t_s = 8.1419 years after 1 January 1969,
or, t_s = 21 February 1977.

4. Termination time for fatalities clustered about line segment L_4 is

t_e = 9.3503 years after 1 January 1969
or, t_e = 7 May 1978.

5. Time length of line segment L_4 is

t_L = 1.2084 years = 1 year and 76 days.

Predicted fatalities under the assumption of arithmetic averaging are shown in Table 5-4. A graph of the line segment L_4, developed under the assumption of arithmetic averaging, also is shown in Figure 5-1.

TABLE 5-4. PREDICTED DATES FOR A-6 EJECTION RELATED FATALITIES, UNDER THE ASSUMPTION OF ARITHMETIC AVERAGING OF TIME

Predicted		Observed	
Date	Fatalities	Date	Fatalities
February 21, 1977	2		
May 22, 1977	1		
September 7, 1977	1		
November 22, 1977	1		
February 7, 1978	1		
April 29, 1978	1		

To summarize, it should be recalled that the first fatality prediction algorithm was deterministically constructed under the following fundamental assumption: time interval between the termination of one cluster of fatalities and the start of a succeeding cluster is governed by logarithmic (or exponential) contraction. The second prediction algorithm was based on the assumption that the above mentioned time interval is constructed on the basis of an arithmetic average of the preceding analogous time intervals. Basic to the entire prediction methodology, whether the assumption is made of logarithmic time contraction or arithmetic averaging, is the fundamental assumption: Ejection conditions in the future will be the same as those in the past. This fundamental assumption cannot be overstressed. An analogous situation is the injury pattern experienced by aircrew who eject from the A-6 aircraft. This problem, addressed at length in Section 6, is one that is analyzed by a goodness-of-fit algorithm. It will be shown that, on the basis of available data and on the basis of invariance of ejection scenarios, future ejection related injury patterns can be predicted with a high degree of confidence.

In retrospect, a non-parametric statistical run test was applied to a set of A-6 fatality data. Temporal randomness (randomness in time) was investigated. It was discovered that A-6 ejection related fatalities do not occur randomly over time. A high degree of confidence can be placed in that assertion. Next, two deterministic prediction schemes were derived: (1) a

logarithmic time interval contraction, and (2) arithmetic time interval averaging. Results are summarized by Tables 5-3 and 5-4, respectively.

Because of the impact and severe nature of making such predictions concerning fatalities, caution must be exercised in interpreting the algorithm. The basic underlying assumption, to render a prediction valid, is that ejection conditions in the future will remain the same as in the past. For example, in this case, the ejection regime will be the same, the flight hour program will be the same, the system functions will be the same, and missions will be the same. Regardless, the mathematics of the prediction scheme is valid, and the fact remains that fatalities did not occur randomly in time upon ejection from A-6 aircraft, hence they can be described quite accurately with deterministic functions. This in itself is sufficient to call attention to A-6 ejections and support a detailed study into the fatalities experienced upon ejection from A-6 aircraft.

5.2 A Statistical Run Test Analysis of A-6 Ejection Related Fatalities Versus Ejection Airspeed

One question pertinent to the A-6 fatality analysis is whether a statistical run test on ejection airspeed would give some insight into the A-6 ejection fatality problem. To answer such a question, consider data in Table 5-5 which lists total A-6 ejection fatalities versus airspeed at which the ejection took place. The particular dichotomy used was that of values above and below the median airspeed. From Table 5-5, it is easy to observe that the median airspeed is 240 knots, so that Table 5-6 is easy to derive. In Table 5-6, a, represents an airspeed above the median, and, b, represents airspeed below the median. The median airspeed is not included in this test.

Here, a run is defined as a sequence of letters of the same kind bounded by letters of the other kind or no letters at all. The null hypothesis to be tested is whether the two samples, set of a's and set of b's, are extracted from the same parent population density function. If the two samples are from the same parent population, the a's and b's will ordinarily be well mixed and the number of runs, u, will be large. If the two populations are widely separated so that their ranges do not overlap, the number of runs will be only 2. In general, differences between the two parent populations will tend to reduce the number of runs, u.

TABLE 5-5. TOTAL A-6 FATALITIES VERSUS EJECTION AIRSPEED IN KNOTS

Date	Fatalities	Ejection Airspeed (knots)
03-07-69	1.	350
03-07-69	2.	350
08-26-69	3.	500
12-26-69	4.	400
12-26-69	5.	400
02-16-70	6.	400
04-22-70	7.	400
06-12-70	8.	460
05-27-72	9.	65
08-09-72	10.	115
10-24-72	11.	40
10-31-72	12.	140
10-31-72	13.	140
09-19-73	14.	400
09-19-73	15.	400
10-09-74	16.	135
11-20-74	17.	100
11-20-74	18.	100
01-13-75	19.	130
06-25-75	20.	240
08-20-75	21.	150

TABLE 5-6. INPUT DATA TO A RUN TEST OF A-6 EJECTION FATALITIES VERSUS AIRSPEED

a, a, a, a, a, a, a, a, b, b,
b, b, b, a, a, b, b, b, b, b

Now formulate the following null hypothesis:

H_o: The sample of a's and the sample of b's (see Table 5-6) are two random samples extracted from the same parent population density function.

From Table 5-6, the following numbers are developed:

$$n_1 = \text{number of a's} = 10$$

$$n_2 = \text{number of b's} = 10$$

$$u = \text{number of runs} = 4$$

To assist with a test of the null hypothesis, use the following exact probability density function:

$$f(u) = \begin{cases} \dfrac{2\dbinom{n_1 - 1}{k - 1}\dbinom{n_2 - 1}{k - 1}}{\dbinom{n_1 + n_2}{n_1}} & ; u = 2k \\[2em] \dfrac{\dbinom{n_1 - 1}{k}\dbinom{n_2 - 1}{k - 1} + \dbinom{n_1 - 1}{k - 1}\dbinom{n_2 - 1}{k}}{\dbinom{n_1 + n_2}{n_1}} & ; u = 2k + 1 \end{cases} \tag{5.21}$$

To test the null hypothesis at the α level of significance, find a positive integer u_o, so that as nearly as possible:

$$\sum_{u=0}^{u_o} f(u) = \alpha . \tag{5.22}$$

Reject the null hypothesis H_o if the observed $u \leq u_o$. In this problem, choose $\alpha = 0.05$.

Using the numbers developed from Table 5-6, equations (5.21) become:

$$f(u) = \begin{cases} \dfrac{2\binom{9}{k-1}^2}{\binom{20}{10}} \; ; \; u = 2k \\[2ex] \dfrac{2\binom{9}{k}\binom{9}{k-1}}{\binom{20}{10}} \; ; \; u = 2k+1 \end{cases} \tag{5.23}$$

Let k = 0 in equation (5.23), then f(0) = f(1) = 0.

Next, let k = 1, then

$$f(2) = \frac{2\binom{9}{0}^2}{\binom{20}{10}} = 0.000010825$$

$$f(3) = \frac{2\binom{9}{1}\binom{9}{0}}{\binom{20}{10}} = 0.0000974257$$

In an entirely analogous fashion, whenever k = 2, 3, ... , 10, values for f(4), f(5); ... ; f(20), f(21) were computed. Complete results are shown in Table 5-7. A graph of results in Table 5-7 is shown in Figure 5-2.

$$\sum_{u=0}^{7} f(u) = 0.0512557$$

Also,

$$\sum_{u=0}^{6} f(u) = 0.0185207 \; .$$

Since

$$\sum_{u=0}^{6} f(u) < \alpha < \sum_{u=0}^{7} f(u) \; ,$$

TABLE 5-7. RESULTS OF COMPUTING EXACT PROBABILITIES FROM THE RUN TEST

k	f(2k)	f(2k + 1)
0	0.000	0.000
1	0.0000108	0.0000974257
2	0.0008768	0.0035073285
3	0.0140293	0.0327350667
4	0.0763818	0.1145727338
5	0.1718591	0.1718591007
6	0.1718591	0.1145727338
7	0.0763818	0.0327350667
8	0.0140293	0.0035073285
9	0.0008768	0.0000974257
10	0.0000108	0.000
TOTALS	0.5263166	0.4736842101

and since $u = 4 < 6 < 7$, reject the null hypothesis at the α level of significance, and conclude that the samples in the run of Table 5-6 are random samples extracted from different parent population density functions.

The null hypothesis could have been stated as follows: H_o: on a chronological time basis, ejection airspeed resulting in a fatality, is a random occurrence. As demonstrated in the preceding analysis this null hypothesis of randomness must be rejected.

It would be interesting to compare these results with the "normal approximation" results. Thus, the mean and variance of the exact probability density functions defined by equation (5.21) are:

$$E(u) = \frac{2\, n_1 n_2}{n_1 + n_2} + 1 \tag{5.24}$$

$$Var(u) = \frac{2\, n_1 n_2 \left(2\, n_1 n_2 - n_1 - n_2\right)}{\left(n_1 + n_2\right)^2 \left(n_1 + n_2 - 1\right)} \tag{5.25}$$

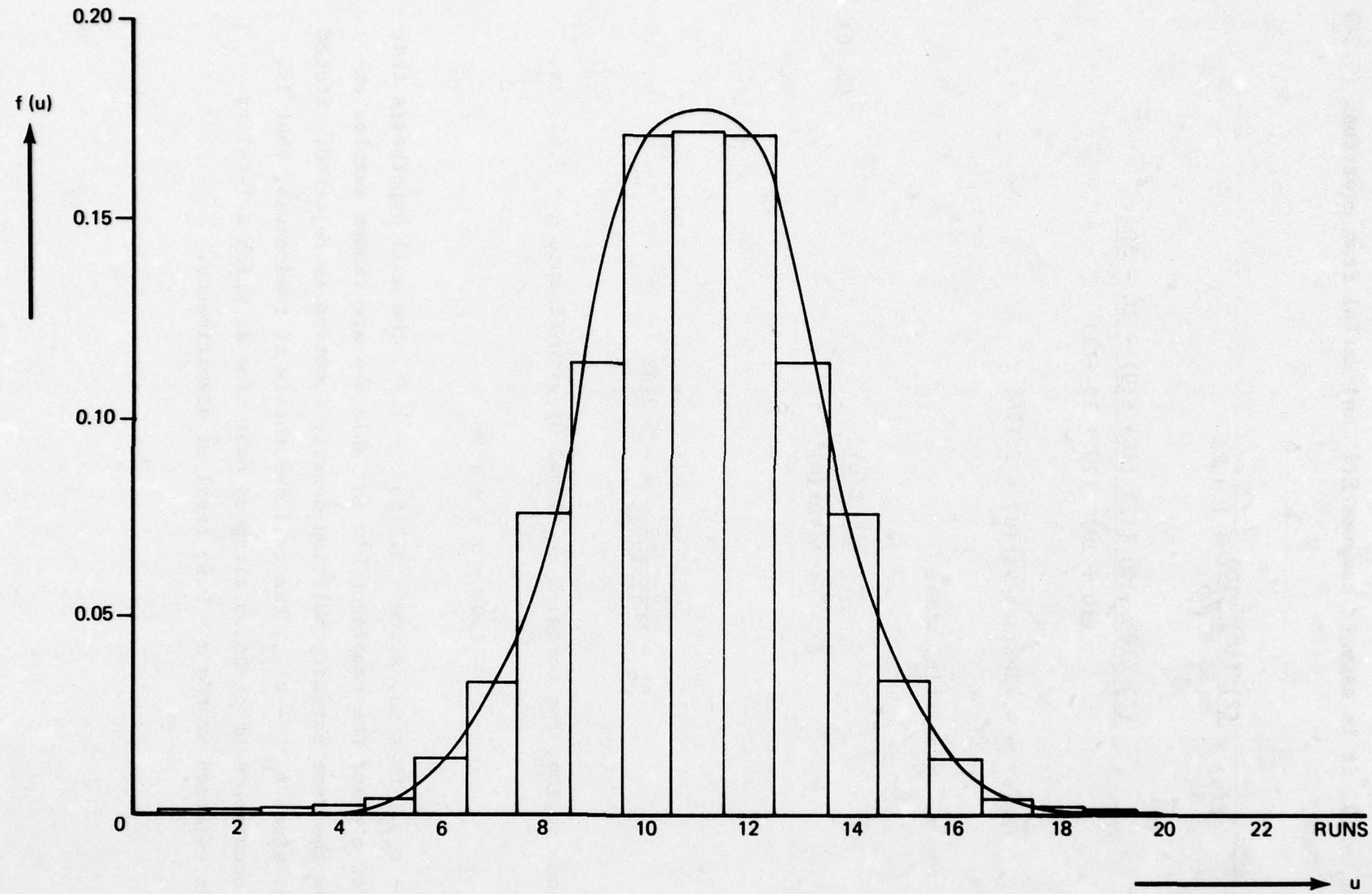

FIGURE 5-2. A FREQUENCY HISTOGRAM OF THE TOTAL NUMBER OF RUNS USING THE EXACT DISCRETE PROBABILITY DENSITY FUNCTION

Since $n_1 = n_2 = 10$, it is easy to compute E(u) and Var(u) from equations (5.24) and (5.25). Thus,

$$E(u) = \frac{(2)\ (10)\ (10)}{(10 + 10)} + 1 = 11$$

$$Var(u) = \frac{(2)\ (10)\ (10)\ [(2)\ (10)\ (10) - 10 - 10]}{(10 + 10)^2\ (10 + 10 - 1)}$$

$$Var(u) = 4.7368; \quad \sqrt{Var(u)} = 2.1764$$

Now compute the z - statistic, where

$$z_o = \frac{u - E(u)}{\sqrt{Var(u)}} \qquad (5.26)$$

Thus,

$$z_o = \frac{4 - 11}{2.17642875} = - 3.2163$$

The acceptance region, for normal z and level of significance $\alpha = 0.05$, is,

$$- 1.96 < z < + 1.96 .$$

Since $z_o < - z_{\alpha/2}$, that is, since $- 3.2163 < - 1.96$, the null hypothesis that the sample of a's and the sample of b's in Table 5-6 are random samples extracted from the same parent population density function is rejected. Stated another way, since $z_o < - z_{\alpha/2}$, the null hypothesis of randomness, that is, the random occurrence of ejection airspeed over time at which a fatality occurred, is rejected at the $\alpha = 0.05$ level of significance.

5.3 Relationship Between A-6 Ejection Velocity and Ejection Related Fatalities

A set of tabulated values of ejection airspeed ranges (knots) varying from 0 through 600 knots was developed from the MORs. Values of number of ejections and number of fatalities, within the noted airspeed ranges, are given in Table 5-8. Percent fatalities, defined as the product of one hundred and the quotient of number of fatalities and number of ejections also is shown. A graph of these data is shown in Figure 5-3. From the graph, it appears that a relatively safe ejection region is from 200 to 400 knots. Numbers on the graph are read as follows: (8/2) = 8 ejections, 2 of which resulted in fatalities.

The point indicated on the graph (Figure 5-3) represents all A-6 ejections and all A-6 ejection related fatalities.

TABLE 5-8. A-6 EJECTION RELATED FATALITIES AS A FUNCTION OF TOTAL EJECTIONS AND EJECTION AIRSPEED IN KNOTS

Ejection Airspeed Range (Knots) ($v_1 \leq v < v_2$)	Ejections	Fatalities	Percent Fatalities
0-100	8	2	25.00
100-200	38	9	23.68
200-300	34	1	2.94
300-400	13	2	15.38
400-500	11	6	54.55
500-600	3	1	33.34
TOTALS:	107	21	19.63

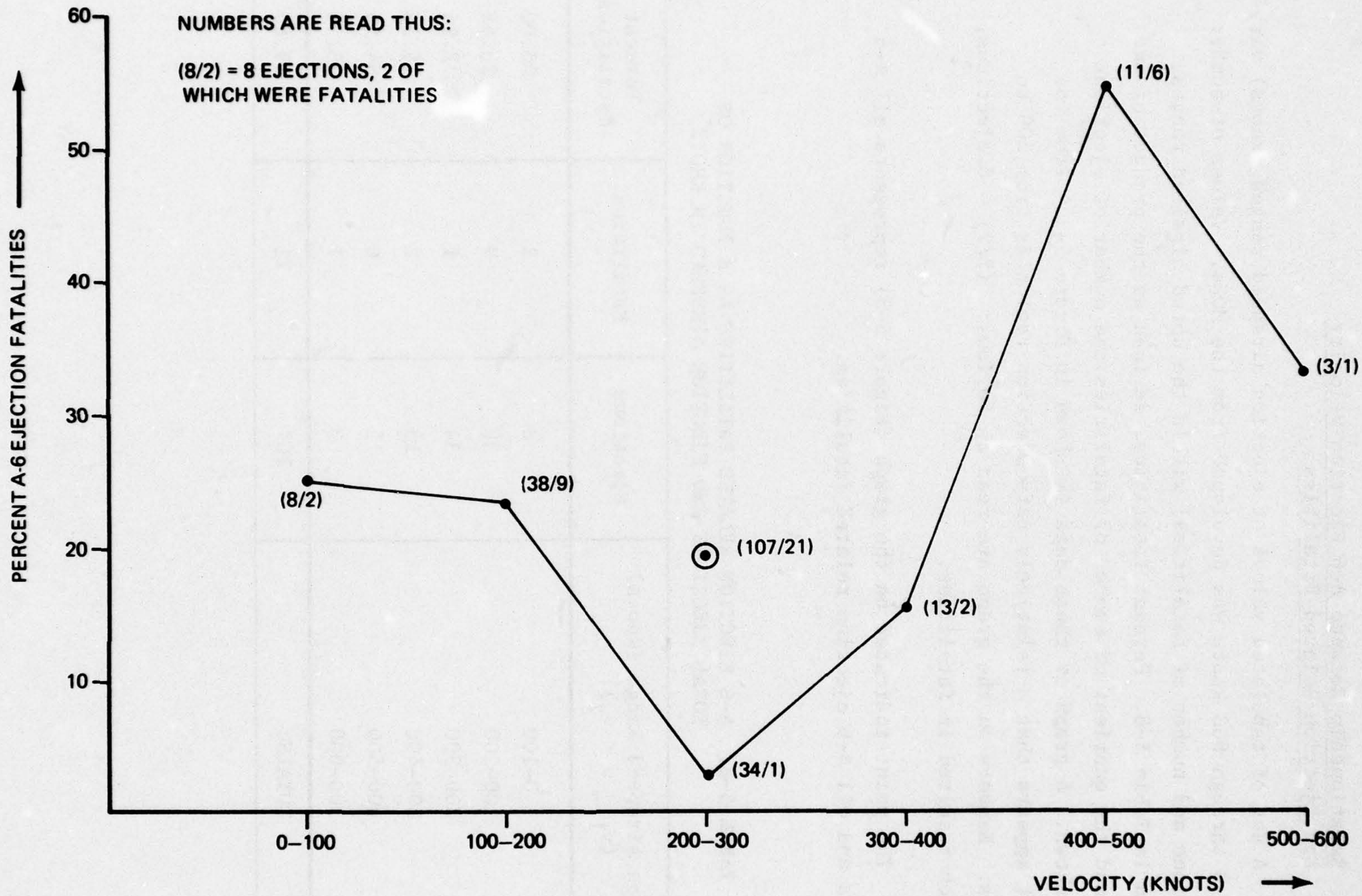

FIGURE 5-3. PERCENT A-6 FATALITIES VERSUS EJECTION VELOCITY

Section 6.0

HYPOTHESIS TESTING

6.0 Hypothesis Testing

Hypothesis testing is a statistical analysis procedure which enables one to do the following: (1) Formulate a hypothesis, (2) Test the hypothesis, and (3) Derive a measure of confidence that can be attached to the results. In restricted circumstances, hypothesis testing can be used as a predictor mechanism.

As defined above, hypothesis testing will be applied to the following: (1) A-4 ejection injury data, (2) the A-6 ejection injury data, and (3) ejection related injury data derived from most U.S. Navy high performance aircraft over the time period 1 January 1969 through 31 December 1975. The hypothesis to be tested is that the ejection injury data is a random sample extracted from a given parent probability density function. In this case, it will be learned that each data set mentioned above is extracted from different gamma parent density functions. The A-4 data, as well as the injury data from seven U.S. Navy aircraft, are data sets extracted from similar gamma parent density functions. The A-6 data set, though extracted from a gamma parent density function is quite different from the other two data sets. Goodness-of-fit of the test is achieved with the Chi-square statistic.

6.1 An Analysis of the A-4 Ejection Related Injury Data

To apply hypothesis testing to actual data, consider the A-4 injury data over the years 1969-1975 as found in the MORs. The following null hypothesis is formulated:

H_o: The A-4 injury data is a random sample of size 241 extracted from a gamma parent density function.

Injury data are summarized in Table 6-1.

The gamma density function is

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta} \; ; \; x \geq 0 \qquad (6.1)$$

Here, let $\alpha = 1$, since a graph of the data (see Figure 6-1) appears exponential in nature. Thus, equation (6.1) becomes:

$$f(x) = a\, e^{-ax} \; ; \; a = \frac{1}{\beta} \; ; \; x \geq 0 \qquad (6.2)$$

TABLE 6-1. INJURY PROFILE FOR A-4 EJECTIONS

Injury Code	Number of Injuries	Percentage of Total Injuries
No/minimal	109	45.2
Minor	60	24.9
Major	35	14.5
Fatal	29	12.0
Lost and Unknown	8	3.3

The numerical procedure for computing the constant a in equation (6.2) is summarized in Table 6-2. A simple arithmetic averaging technique was used to get a = 0.46208.

To compute expected frequency, equation (6.2) is integrated between definite limits. Thus,

$$\left.\begin{aligned} F_i &= P(A < x < B) = \int_A^B a\, e^{-ax}\, dx \\ F_i &= P(A < x < B) = e^{-aA} - e^{-aB} \end{aligned}\right\} \quad (6.3)$$

The Chi-square test with 4 degrees of freedom can now be computed.

$$\chi^2_{c/d.f.=4} = \sum_{i=1}^{5} \frac{(f_i - F_i)^2}{F_i}\ , \quad (6.4)$$

where f_i is the observed frequency of injury occurrence, and F_i is the expected frequency.

From column 8 in Table 6-2, $\chi^2_{c/d.f.=4} = 0.0385497$, or translated to the 241 ejection injury base,

$$\chi^2_{c/d.f.=4} = 9.2904$$

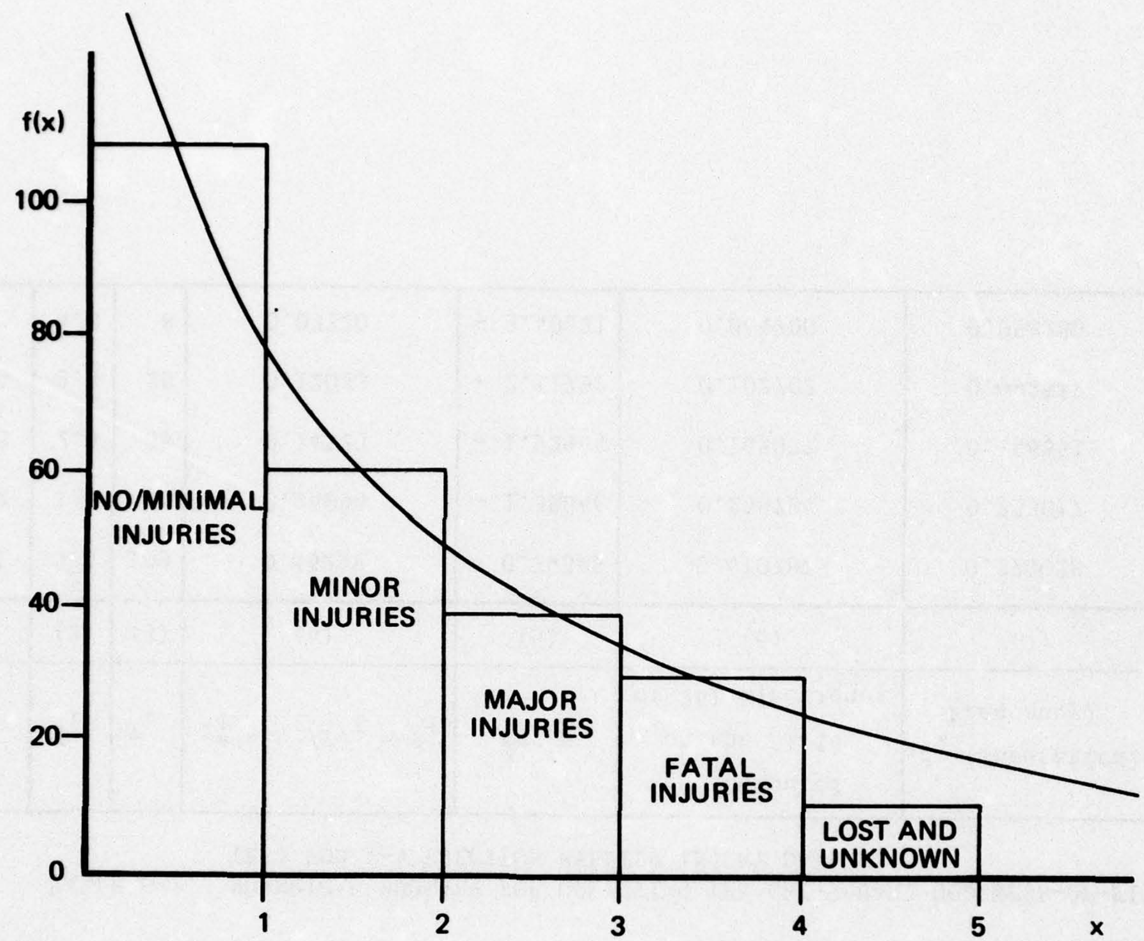

FIGURE 6-1. GRAPH OF A-4 INJURY PROFILE FROM MORs DATA

TABLE 6-2. NUMERICAL SUMMARY FOR COMPUTING THE CHI-SQUARE GOODNESS-OF-FIT TEST FOR A-4 EJECTION RELATED INJURY DATA

Injury Code		x_i	y_i	$y_i^* = y_i/\Sigma y_i = f_i$	$\ell n\ y_i^*$	y_i^* (Computed on the Basis of 241 Ejections)	F_i (Theoretical Frequency)	$(f_i - F_i)^2/F_i$
(1)		(2)	(3)	(4)	(5)	(6)	(7)	(8)
G	1	0.5	109	0.45228	- 0.79345	0.410787	0.370028	0.018283
F	2	1.5	60	0.24896	- 1.39046	0.258784	0.233017	0.001090
B	3	2.5	35	0.14523	- 1.92944	0.163027	0.146851	0.000018
A	4	3.5	29	0.12033	- 2.11752	0.102702	0.092512	0.008365
L & U	5	4.5	8	0.03320	- 3.40521	0.064700	0.058280	0.010793

The tabulated value at the significance level $\alpha = 0.05$ is

$$\chi^2_{T/d.f.=4;\ \alpha=0.05} = 9.488$$

Since $\chi_c^2 < \chi_T^2$, the null hypothesis is accepted and it can be stated with 95 percent confidence that this particular data sample of A-4 injuries was extracted from a parent density function which is gamma distributed.

This is one example of use of hypothesis testing. A great many additional hypotheses, similar to the above can be formulated from the MORs data. Consider next the A-6 ejection injury data.

6.2 An Analysis of the A-6 Ejection Related Injury Data

A graph of the A-6 ejection related injury data is shown in Figure 6-2. Inspection of Figure 6-2 leads to the following null hypothesis:

> H_o: The A-6 ejection related injury data is a random sample extracted from a gamma parent probability density function.

To test validity of this hypothesis, first compute α, and β in the gamma parent probability density function:

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta} \ ; \ x \geq 0 \ , \tag{6.5}$$

then compute theoretical frequency using equation (6.5), finally compute the Chi-square statistic, and compare the computed value with the value given at the level of significance, $\alpha = 0.05$ in this case. Numbers used to compute α and β, in equation (6.5), are developed from Table 6-3.

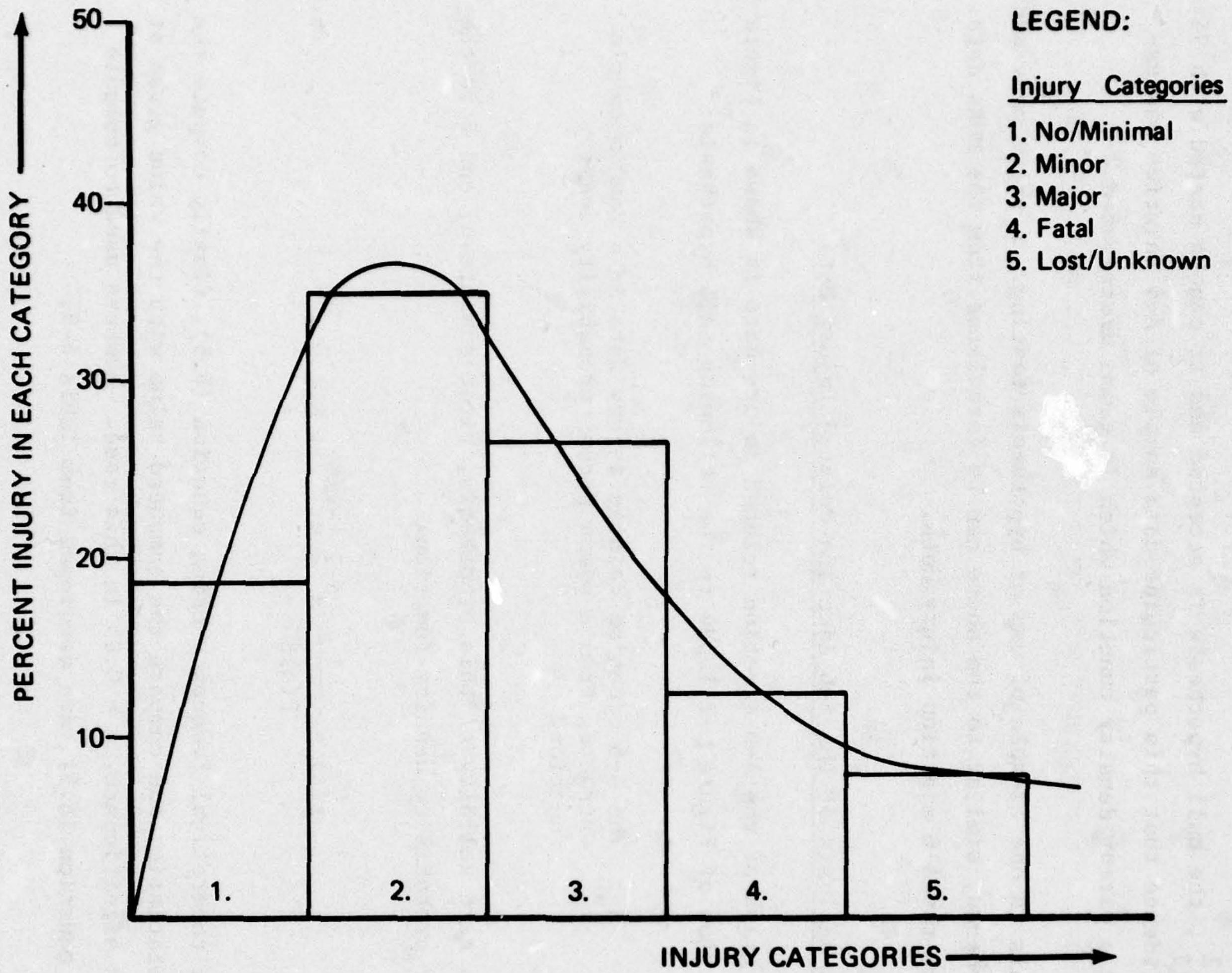

FIGURE 6-2. HISTOGRAM OF A-6 EJECTION RELATED INJURY DATA

TABLE 6-3. NUMBERS DEVELOPED FROM A-6 INJURY DATA

x	f(x)	f*(x)	ℓn x	ℓn f*(x)
0.5	20	0.1869	- 0.693147	- 1.677182
1.5	38	0.3552	0.405465	- 1.035074
2.5	27	0.2523	0.916291	- 1.377136
3.5	15	0.1402	1.252763	- 1.964685
4.5	7	0.0654	1.504077	- 2.727233
TOTAL:	107	1.0000		

Equation (6.5) can be written as follows:

$$w = A_o + A_1 x_1 + A_2 x_2 \ , \tag{6.6}$$

where

$$\left.\begin{aligned} w &= \ell n\ f^*(x)\ ; & x_1 &= \ell n\ x \\ A_o &= \ell n\ [1/\Gamma(\alpha)\beta^{\alpha}]\ ; & A_2 &= -\ 1/\beta \\ A_1 &= \alpha - 1\ ; & x_2 &= x \end{aligned}\right\} \tag{6.7}$$

Data from Table 6-3 permits the following equations to be written:

$$\left.\begin{aligned} -\ 1.677182 &= A_o - 0.693147\ A_1 + 0.5\ A_2 \\ -\ 1.035074 &= A_o + 0.405465\ A_1 + 1.5\ A_2 \\ -\ 1.377136 &= A_o + 0.916291\ A_1 + 2.5\ A_2 \\ -\ 1.964685 &= A_o + 1.252763\ A_1 + 3.5\ A_2 \\ -\ 2.727233 &= A_o + 1.504077\ A_1 + 4.5\ A_2 \end{aligned}\right\} \tag{6.8}$$

Average the first three of equations (6.8), then average the next three, and finally average the last three to get the following three linear equations:

$$\left.\begin{aligned} -\ 1.363131 &= A_o + 0.209536\ A_1 + 1.5\ A_2 \\ -\ 1.458965 &= A_o + 0.858173\ A_1 + 2.5\ A_2 \\ -\ 2.023018 &= A_o + 1.224377\ A_1 + 3.5\ A_2 \end{aligned}\right\} \tag{6.9}$$

Solving equations (6.9) simultaneously yields

$$\left.\begin{aligned} A_o &= 0.046223 \\ A_1 &= 1.65781 \\ A_2 &= -1.17115 \end{aligned}\right\} \qquad (6.10)$$

From equations (6.7) these constants become:

$$\left.\begin{aligned} \alpha &= 2.65781 \\ \beta &= 0.85386 \\ K &= [1/\Gamma(\alpha)\ \beta^{\alpha}] = 1.047308 \end{aligned}\right\} \qquad (6.11)$$

From equations (6.11), it is seen that

$$\Gamma(\alpha) = 1.453064 \qquad (6.12)$$

Now $\Gamma(2.65781) = 1.65781\ \Gamma(1.65781)$, hence from tabulated values of the gamma function

$$\Gamma(2.65781) = 1.494234 \qquad (6.13)$$

A mean value of $\Gamma(\alpha)$ is

$$\left.\begin{aligned} \bar{\Gamma}(\alpha) &= (1.453064 + 1.494234)/2 \\ \bar{\Gamma}(\alpha) &= 1.473649 \end{aligned}\right\} \qquad (6.14)$$

Using $\bar{\Gamma}(\alpha)$, equations (6.11) become:

$$\left.\begin{aligned} \alpha &= 2.65781 \\ \beta &= 0.85386 \\ K &= [1/\bar{\Gamma}(\alpha)\ \beta^{\alpha}] = 1.032683275 \end{aligned}\right\} \qquad (6.15)$$

Equation (6.5) can now be written thus:

$$f(x) = 1.032683\ x^{1.65781}\ e^{-1.17115x} \qquad (6.16)$$

Equation (6.16) is the basic equation from which theoretical frequencies can be obtained.

Another method for obtaining the parameters α and β is from the moment generating function from which it is found that for the gamma density function:

$$\left.\begin{aligned} \mu &= \beta(\alpha + 1) \\ \sigma^2 &= \beta^2(\alpha + 1) \end{aligned}\right\} \qquad (6.17)$$

From the data in Table 6-3, estimators for α and β are:

$$\left.\begin{aligned} \hat{\alpha} &= \frac{\bar{x}^2 - s^2}{s^2} = 2.190075 \\ \hat{\beta} &= \frac{s^2}{\bar{x}} = 0.640128 \ , \end{aligned}\right\} \qquad (6.18)$$

where

$$\left.\begin{aligned} \bar{x} &= 2.042056 \\ s^2 &= 1.307177 \end{aligned}\right\} \qquad (6.19)$$

Two methods were used to compute theoretical frequency from equation (6.16): (1) a power series representation, thus,

$$P(a < x < b) = \frac{1}{\Gamma(\alpha)} \sum_{n=0}^{\infty} \frac{(-1)^n \ (b^{\alpha+n} - a^{\alpha+n})}{(\alpha + n) \ \beta^{\alpha+n} \ n!} \qquad (6.20)$$

and (2) a numerical integration scheme using the trapezoidal rule. An example of application of the trapezoidal rule is shown in Table 6-4. Results from both integration schemes are found in Table 6-5.

TABLE 6-4. NUMERICAL INTEGRATION RESULTS USING THE TRAPEZOIDAL RULE

x	$f_i^*(x)$	$[f_i^*(x_{n-1}) + f_i^*(x_n)]/20$
2.0	0.3132048	---
2.1	0.3020605	0.030763
2.2	0.2902184	0.029614
2.3	0.2778861	0.028405
2.4	0.2652445	0.027157
2.5	0.2524497	0.025885
2.6	0.2396351	0.024604
2.7	0.2269135	0.023327
2.8	0.2143788	0.022065
2.9	0.2021084	0.020824
3.0	0.1901649	0.019614
	TOTAL:	Σ = 0.252258

From tabulated values of the Chi-square statistic,

$$\chi^2_{T/d.f.=3;\ \alpha=0.05} = 7.81$$

The numerical value of the Chi-square statistic as computed by the power series expression is

$$\chi^2_{c/d.f.=3;\ \alpha=0.05} = (107)(0.0076150)$$

$$\chi^2_{c/d.f.=3;\ \alpha=0.05} = 0.814805 \text{ (Power Series)}$$

From the Trapezoidal Rule,

$$\chi^2_{c/d.f.=3;\ \alpha=0.05} = 0.243429$$

TABLE 6-5. NUMERICAL RESULTS FROM AN ANALYSIS OF A-6 EJECTION RELATED INJURY DATA

x	$f_i^*(x)$ (Observed)	$f^*(x)$ (From Equation 6.16)	$F_i = \int_a^b f(x)\,dx$ (Power Series)	$F_i = \int_a^b f(x)\,dx$ (Trapezoidal Rule)	$\chi_i^2 = (f_i - F_i)^2/F_i$ (Power Series)	$\chi_i^2 = (f_i - F_i)^2/F_i$ (Trapezoidal Rule)
0.5	0.1869	0.1822	0.171962	0.171257	0.0012976	0.001429
1.5	0.3552	0.3491	0.340685	0.338463	0.0006184	0.0008276
2.5	0.2523	0.2524	0.249018	0.252257	0.00004326	0.0000000073
3.5	0.1402	0.1367	0.115278	0.138722	0.0053879	0.0000157
4.5	0.0654	0.0643	0.069721	0.065824	0.00026780	0.000002731
TOTALS:	1.0000	0.9847	0.946664	0.966523	0.0076150	0.002275

Since χ_c^2, as computed either by the power series representation or the trapezoidal rule is less than χ_T^2, the null hypothesis, H_o, is accepted and we conclude that with 95 percent confidence the random sample of A-6 ejection injury data was extracted from a gamma parent probability density function.

6.3 An Analysis of Ejection Related Injury Data from Seven U.S. Navy High Performance Aircraft

For comparison of A-6 injury data with injury data from several U.S. Navy high performance aircraft, an injury frequency histogram was constructed, and an exponential parent density function hypothesized. A graphical display is shown in Figure 6-3. Raw data, on which the analysis is based, are shown in Tables 6-6 and 6-7.

The following null hypothesis was formulated:

H_o: The sample of injury data from seven U.S. Navy high performance aircraft is a random sample extracted from a parent probability density function which is exponentially distributed.

The parent probability density function, hypothesized in the null hypothesis, is of the form:

$$f(x) = a\, e^{-ax}; \; x \geq 0; \; a > 0 \tag{6.21}$$

To determine the parameter a, in the parent probability density function, data in Table 6-8 were used. From equation (6.21) the following relationship is easy to obtain:

$$w = A_o - A_1 x_1 \tag{6.22}$$

where

$$\left.\begin{aligned} w &= \ln f^*(x), \; A_o = \ln a \\ A_1 &= a \end{aligned}\right\} \tag{6.23}$$

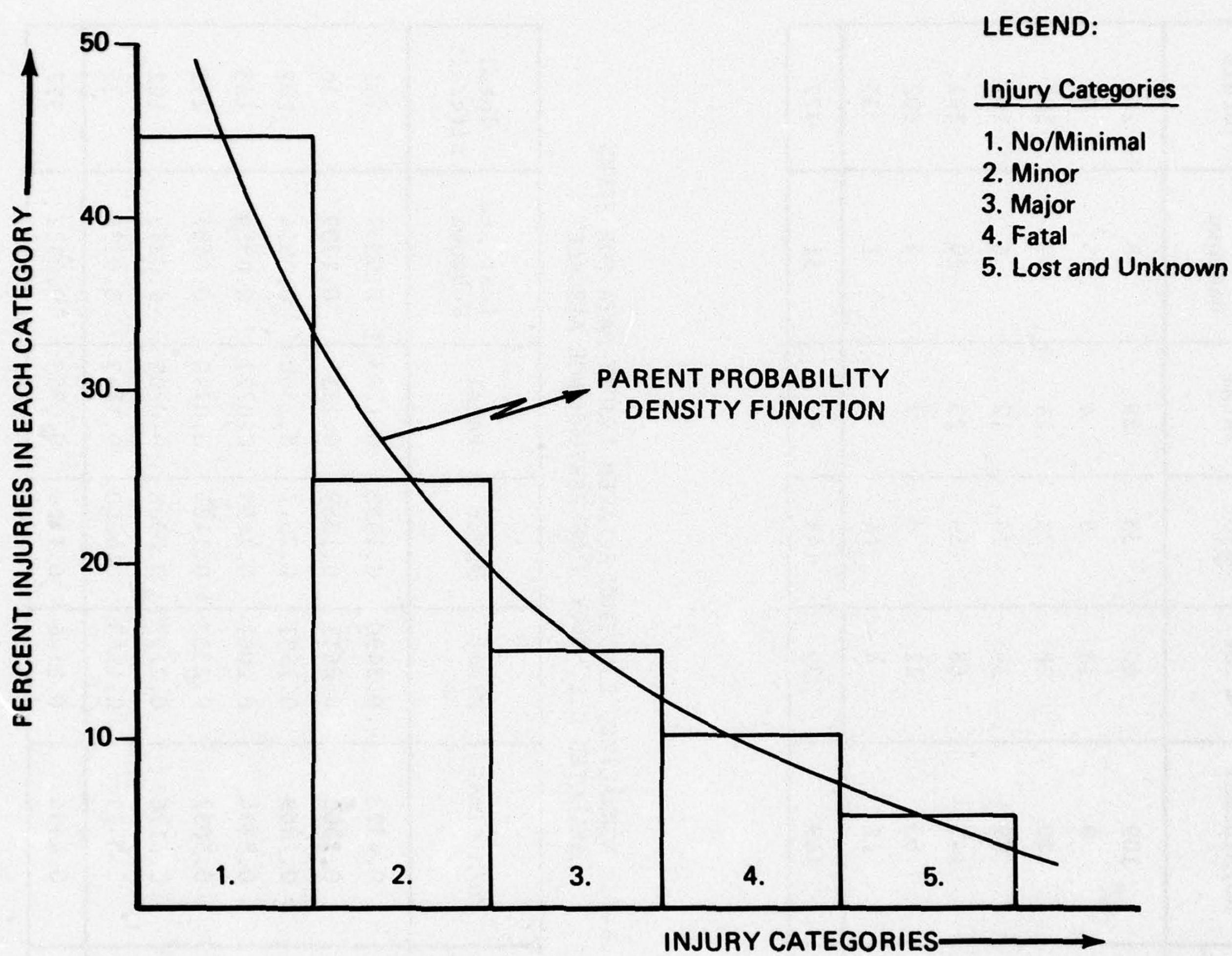

FIGURE 6-3. HISTOGRAM OF EJECTION RELATED INJURY DATA FROM SEVEN U.S. NAVY HIGH PERFORMANCE AIRCRAFT

TABLE 6-6. EJECTION RELATED INJURY DATA FOR SEVEN SELECTED U.S. NAVY HIGH PERFORMANCE AIRCRAFT

Injury → / Aircraft ↓	No/Minimal	Minor	Major	Fatal	Lost and Unknown	Totals
A-4	109	60	35	29	8	241
A-5	9	13	5	4	5	36
A-6	20	38	28	14	7	107
A-7	89	34	24	12	6	165
F-4	147	68	33	23	20	291
F-8	63	22	6	7	4	102
F-9	12	4	14	4	1	35
TOTALS:	449	239	145	93	51	977

TABLE 6-7. NORMALIZED EJECTION RELATED INJURY DATA FOR SEVEN SELECTED U.S. NAVY HIGH PERFORMANCE AIRCRAFT

Injury → / Aircraft ↓	No/Minimal	Minor	Major	Fatal	Lost and Unknown	Total Aircraft
A-4	0.4523	0.2490	0.1452	0.1203	0.0332	241
A-5	0.2500	0.3611	0.1389	0.1111	0.1389	36
A-6	0.1869	0.3551	0.2617	0.1308	0.0654	107
A-7	0.5394	0.2061	0.1455	0.0727	0.0364	165
F-4	0.5052	0.2337	0.1134	0.0790	0.0687	291
F-8	0.6176	0.2157	0.0588	0.0686	0.0392	102
F-9	0.3429	0.1143	0.4000	0.1143	0.0286	35
All A/C	0.4596	0.2446	0.1484	0.0952	0.0522	977

TABLE 6-8. DATA USED TO COMPUTE THE PARAMETER IN AN EXPONENTIAL PARENT PROBABILITY DENSITY FUNCTION FROM WHICH THE RANDOM SAMPLE OF EJECTION RELATED INJURY DATA FOR SEVEN AIRCRAFT WAS EXTRACTED

x	f(x)	$f_i^*(x)$	$\ell n\ f_i^*(x)$
0.5	449	0.4596	- 0.777399
1.5	239	0.2446	- 1.408131
2.5	145	0.1484	- 1.907844
3.5	93	0.0952	- 2.351775
4.5	51	0.0522	- 2.952673
TOTALS:	977	1.0000	---

The following linear equations are an immediate consequence of Table 6-8 and equation (6.22):

$$\left.\begin{aligned} -0.777399 &= A_o - 0.5a \\ -1.408131 &= A_o - 1.5a \\ -1.907844 &= A_o - 2.5a \\ -2.351775 &= A_o - 3.5a \\ -2.952673 &= A_o - 4.5a \end{aligned}\right\} \tag{6.24}$$

Get an arithmetic average of the first three equations in (6.24) then the last three equations. Thus:

$$\left.\begin{aligned} -1.364458 &= A_o - 1.5a \\ -2.404017 &= A_o - 3.5a \end{aligned}\right\} \tag{6.25}$$

Solve these simultaneously to get:

$$a_1 = 0.5198195$$

$$A_o = \ln a = - 0.58472875 \rightarrow a_2 = 0.557256997$$

Let $\bar{a} = a = (a_1 + a_2)/2$

$$a = 0.5385382 \tag{6.26}$$

Thus, equation (6.21) is written:

$$f(x) = 0.5385382\ e^{-0.5385382x} \tag{6.27}$$

To get theoretical frequencies over the designated frequency cells, recall that

$$P(A < x < B) = \int_{x=A}^{B} f(x)\ dx \tag{6.28}$$

Substitute equation (6.21) into (6.28) to get:

$$P(A < x < B) = e^{-aA} - e^{-aB} \tag{6.29}$$

Recall that

$$P(A < x < B) = P(0 < x < B) - P(0 < x < A),$$

then from equation (6.29),

$$P(0 < x < B) = 1 - e^{-aB},$$

or for the numbers being used:

$$P(0 < x < B) = 1 - e^{-0.5385382B} \tag{6.30}$$

Theoretical frequencies are computed from equation (6.30). Numerical values are given in Table 6-9.

TABLE 6-9. NUMERICAL VALUES OF OBSERVED FREQUENCY, THEORETICAL FREQUENCY, AND THE CHI-SQUARE VALUES

Interval Frequency Cell	Observed Frequency $= f_i$	Theoretical Frequency $= F_i$	Chi-Square Values $= (f_i - F_i)^2/F_i$
$0 \leq x < 1$	0.4596	0.416399	0.004482
$1 \leq x < 2$	0.2446	0.243011	0.000010
$2 \leq x < 3$	0.1484	0.141821	0.000305
$3 \leq x < 4$	0.0952	0.082767	0.001869
$4 \leq x \leq 5$	0.0522	0.048303	0.000314
TOTALS:	1.0000	0.932301	0.006981

Translated to the sample size of 977 aircraft,

$$\chi_c^2 = (977)(0.006981) = 6.820449.$$

The theoretical value of the Chi-square statistic at $\alpha = 0.05$ and 4 degrees of freedom is

$$\chi^2_{T/d.f.=4;\ \alpha=0.05} = 9.49.$$

Since $\chi_c^2 < \chi_T^2$, accept the null hypothesis and conclude that at the $\alpha = 0.05$ level of significance the sample was a random sample of size 977 extracted from a parent probability density function which is exponentially distributed.

A comparison between the A-6 ejection related injury pattern and the ejection related injury pattern for seven U.S. Navy high performance aircraft is shown in Figure 6-4.

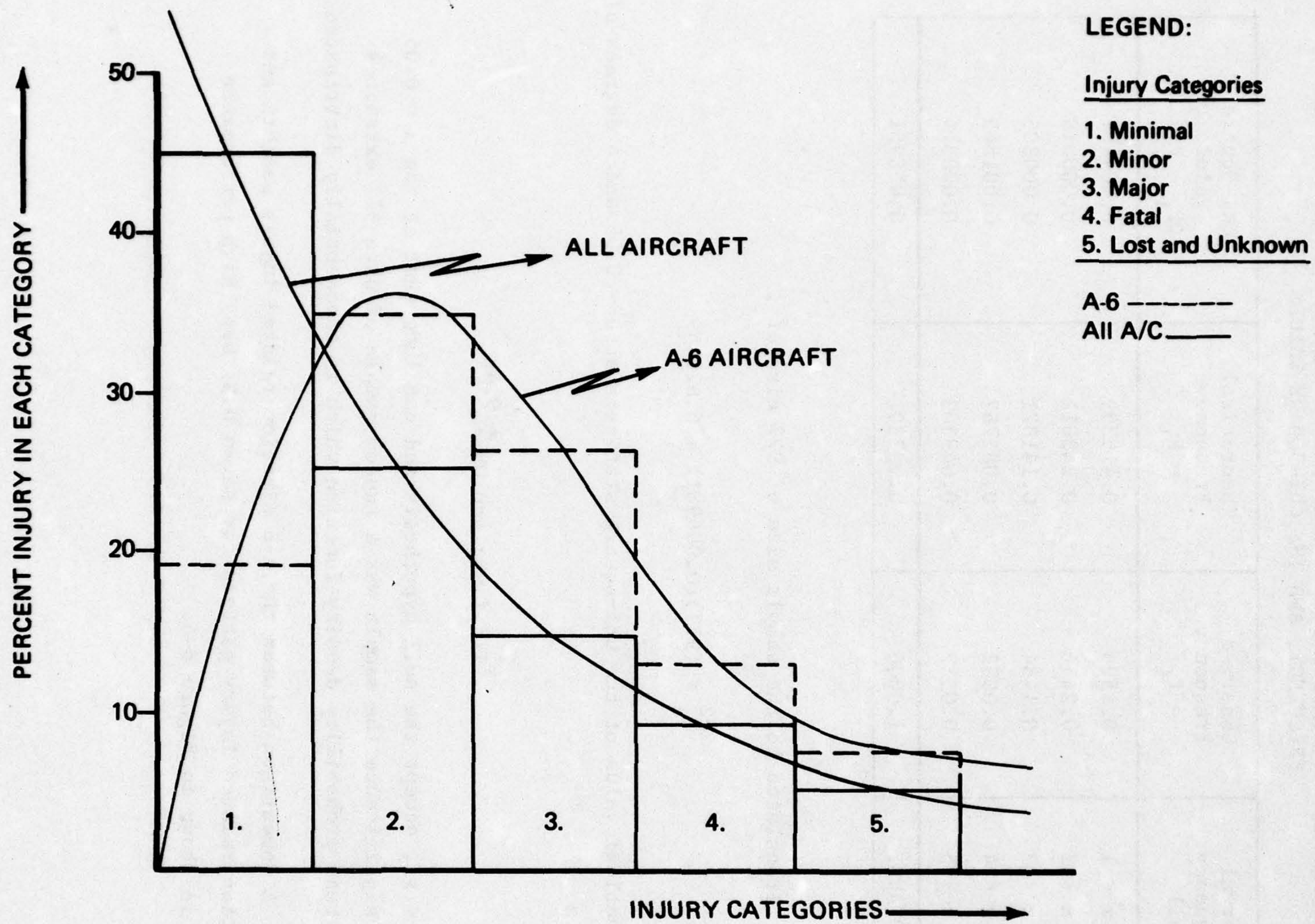

FIGURE 6-4. COMPARISON OF THE A-6 EJECTION INJURY PATTERN WITH THE EJECTION INJURY PATTERN FOR SEVEN AIRCRAFT

Section 7.0

SUMMARY

7.0 Summary

This report contains the results of statistical techniques applied to ejection related injury data which were developed as a consequence of ejections from U.S. Navy high performance aircraft. Data, taken from the MORs, were collected over the time period 1 January 1969 through 31 December 1975. The report was derived over the time period 15 May 1976 through 15 August 1976.

The statistical tests applied to ejection related injury data consisted of the following:

- Two-Way Analysis of Variance
- Non-Parametric Trend Test
- Non-Parametric Run Test
- Deterministic Analyses
  - Numerical Techniques
  - Deterministic Prediction
- Hypothesis Testing using the following parent probability density functions:
  - Runs Discrete Density Function
  - Normal Density Function
  - Gamma Density Function
  - Exponential Density Function
- Chi-Square Density Function was used to test Goodness-of-Fit of random samples to the preceding parent probability density functions.

A summary of tests performed, results obtained, and confidence in the results is found in Table 7-1. There it is noted that underlying trends are strongly suspected to occur in ejection injury data developed from the following U.S. Navy high performance aircraft: A-4, A-6, A-7, F-4, and F-8. There is no reason to suspect a trend in the A-5 and F-9 ejection related injury data.

Use of the run test revealed that on a chronological basis, fatalities occur randomly upon ejection from the A-4, A-7, F-4, and F-8 aircraft, but they do not occur randomly upon ejection from the A-6.

TABLE 7-1. SUMMARY OF SELECTED STATISTICAL ANALYSIS RESULTS

HYPOTHESES TO BE TESTED			STATISTICAL TEST USED	INFERENCE	CONFIDENCE IN RESULTS
1.	On a chronological time basis, fatalities occur randomly upon ejection from A-6 aircraft.		Run Test	Reject this Hypothesis	95 Percent
2.	On a chronological time basis, fatalities occur randomly upon ejection from A-4, A-7, F-4 and F-8 aircraft.		Run Test	Accept this Hypothesis	95 Percent
3.	On a time basis, the A-4, A-6, A-7, F-4, and F-8 ejection related injury data do not contain an underlying trend.		Trend Test	Reject this Hypothesis	95 Percent
4.	On a time basis, the A-5 and F-9 ejection related injury data do not contain an underlying trend.		Trend Test	Accept this Hypothesis	95 Percent
5.	The A-4 ejection related injury data is a random sample which is extracted from a known parent probability density function (Exponential).		Chi-Square Test	Accept this Hypothesis	95 Percent
6.	The A-6 ejection related injury data is a random sample which is extracted from a known parent probability density function (Gamma).		Chi-Square Test	Accept this Hypothesis	95 Percent
7.	Ejection related injury data from seven U.S. Navy high performance aircraft the A-4, A-5, A-6, A-7, F-4, F-8, and F-9, considered collectively, is a random sample which is extracted from a known parent probability density function (Exponential).		Chi-Square Test	Accept this Hypothesis	95 Percent
8.	8.1	There is no difference among the average number of ejection fatalities caused by hardware hazards, aircrew judgment, and environmental conditions.	Two-Way Analysis of Variance Test, one observation per cell.	Reject this Hypothesis	95 Percent
	8.2	There is no difference among the average number of ejection fatalities associated with ejection from the A-4, A-7, A-6, F-4, and F-8 aircraft.		Reject this Hypothesis	95 Percent
9.	9.1	There is no difference among the average number of ejection fatalities caused by hardware hazards, aircrew judgment, and environmental conditions.	Two-Way Analysis of Variance Test, one observation per cell.	Accept this Hypothesis	95 Percent
	9.2	There is no difference among the average number of ejection fatalities caused upon ejection with either the ESCAPAC or Martin-Baker ejection seats.		Accept this Hypothesis	95 Percent

Parent probability density functions were discovered for ejection related injury data from the A-6 aircraft (a generalized gamma density function), the A-4 (an exponential density function), and seven high performance aircraft, the A-4, A-5, A-6, A-7, F-4, F-8, and F-9, considered collectively as one sample (the exponential density function). The Chi-square statistic was used to test goodness-of-fit of the sample data to the parent density functions hypothesized.

Analysis of variance techniques were used to detect differences among various ejection related fatality scenarios. Components of the scenarios were: hardware hazards, aircrew judgment, environmental conditions, high performance aircraft (A-4, A-7, A-6, F-4, and F-8), and ejection seats (ESCAPAC and Martin-Baker).

Statistical investigations conducted to date, while informative, should be categorized as initial investigations only. Certain problem areas have been detected. It is believed that certain other problem areas will be discovered as the investigation continues. Areas which need immediate pursuing are the following:

- An in-depth investigation of all A-6 ejections.
- Determine the underlying injury trends in the A-4, A-6, A-7, F-4, and F-8 ejection related injury data.
- Reconcile any apparently inconsistent results detected by the analysis of variance techniques.
- After sufficient analytical results have been obtained from both preliminary and in-depth investigations, the question of reasons for the particular ejection injury scenario needs to be addressed.
- Statistical and deterministic techniques should continue to be applied to additional ejection data, so that predictive trends can be identified in ejection equipment failure data as well as ejection related injury data.

APPENDIX A

LIST OF SYMBOLS

APPENDIX A

LIST OF SYMBOLS

A.1 English Symbols

A - lower limit in a probability calculation; thus, $P(A \leq x \leq B)$ reads the probability that x lies between A and B; here A is the least value that x can assume.

- injury designation in the Medical Officer's Reports indicating a fatality.

- ejection related fatality which is attributed primarily to the aircrew.

a - a numerical value of a data point which exceeds the median of the data set.

A_k - constants in a polynomial representation of $\bar{T}$ versus $\bar{t}$.

A_0, A_1 - constants used in regression analysis to determine the parameter in an exponential parent probability density function.

A_0, A_1, A_2 - constants used in regression analysis to determine the two parameters in a gamma parent probability density function.

ANOVA - Analysis of Variance.

b - parameter in an exponential parent probability density function.

B - upper limit in a probability calculation. Thus, $P(A \leq x \leq B)$, read the probability that x lies between A and B, here B is the greatest value that x can assume.

- injury designation in the Medical Officer's Reports, indicating a major injury.

B_1, B_2 - constants to be determined in an exponential equation (5.10).

b_j - slope of the j^{th} line segment L_j.

c - designates columns in an Analysis of Variance computation.

c_1, c_2 - constants to be determined in an exponential equation (5.18).

C_j - sum of all the row elements in the j^{th} column of an Analysis of Variance matrix.

D_1, D_2 - differences in iterative schemes, for two separate difference equations which approach zero as the number of iterations increases.

E - ejection related fatality which can be attributed primarily to the environment.

e - base of the natural logarithms.

E(u) - mean of the parent probability density function for the random variable u which represents the number of runs, thus:

$$u = 0, 1, 2, \ldots, n.$$

F - injury designation in the Medical Officer's Reports, indicating a minor injury.

F_i - theoretical frequency in the i^{th} frequency cell.

f_i - observed frequency in the i^{th} frequency cell.

f(u) - discrete parent density function of the number of runs u in a non-parametric run test.

$f(x)$	- a dependent variable, may be a function of a discrete or continuous real variable x.
$F_{o,c}$ (m,n)	- theoretical value of the F-statistic with m and n degrees of freedom. The subscript o,c means the test is being performed on the null hypothesis with respect to ANOVA columns.
$\hat{F}_{o,c}$ (m,n)	- value of the F-statistic computed from data, with m and n degrees of freedom; used when testing a null hypothesis with respect to columns in an ANOVA matrix.
$F_{o,r}$ (m,n)	- theoretical value of the F-statistic with m and n degrees of freedom; used when testing a null hypothesis with respect to rows in an ANOVA matrix.
$\hat{F}_{o,r}$ (m,n)	- value of the F-statistic computed from data, with m and n degrees of freedom; used when testing a null hypothesis with respect to rows in an ANOVA matrix.
G	- sum of all the elements in an Analysis of Variance matrix.
	- injury designation in the Medical Officer's Reports indicating no injury or minimal injury.
H	- ejection related fatality which can be attributed primarily to hardware.
H_o	- a null hypothesis to be tested.
$H_{o,c}$	- null hypothesis with respect to the columns in an ANOVA matrix.
$H_{o,r}$	- null hypothesis with respect to the rows in an ANOVA matrix.
I	- an integer representation in a computer flow diagram; so also is J, K, L, M, and N.

i - index of summation; can be used as a discrete counting variable.

j - index of summation; can be used as a discrete counting variable.

K - a constant defined as follows: $K = [\beta^{\alpha} \Gamma(\alpha)]^{-1}$; used to determine the parameters in a gamma parent probability density function.

k - an index of summation.

- a parameter in the discrete parent probability density function of the number of runs in a non-parametric run test. Thus, $u = 2k$ if u is an even number of runs, and $u = 2k + 1$ if u is an odd number of runs.

L - injury designation in the Medical Officer's Reports; indicates an aircrewman that was lost, that is, his body was not recovered.

L_j - represents a line segment over the j^{th} time interval; $j = 1, 2, 3, 4$.

m_j - slope of the j^{th} line segment L_j.

MORs - Medical Officer's Reports.

N - number of data elements in the sample under a trend test investigation.

N_1 - lower limit for the number of trends in the null hypothesis acceptance region of a non-parametric trend test.

N_2 - upper limit for the number of trends in the null hypothesis acceptance region of a non-parametric trend test.

N_c - number of trends computed in a non-parametric trend test. The subscript c represents computed.

n - index of summation.

n_1 - the sum of all elements of a given designation in a dichotomous (0, 1; or a, b) non-parametric run test; say the sum of all the 1's or all the a's.

n_2 - the sum of all elements of a given designation in a dichotomous (0, 1; or a, b) non-parametric run test; say the sum of all the 0's or all the b's.

n_j - total number of data points in the j^{th} time interval.

$P(A \leq x \leq B)$ - the probability that x lies between the lower limit A and the upper limit B.

r - designates rows in an Analysis of Variance computation.

R_i - sum of all the column elements in the i^{th} row of an Analysis of Variance matrix.

s^2 - an unbiased estimate of population variance defined thus:

$$s^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n - 1}$$

SAR - Search And Rescue.

SSC - sum of squares on the column elements in an Analysis of Variance matrix.

SSE - sum of the squares of the errors in an Analysis of Variance matrix.

SSR - sum of squares on the row elements in an Analysis of Variance matrix.

SST - total sum of squares of the elements in an Analysis of Variance matrix.

T_i - total fatalities at the i^{th} data point.

$\bar{T}_j$ - arithmetic mean of the number of fatalities in the j^{th} time interval; sometimes written as $\bar{T}_{L_j}$.

t_i - time coordinate of the i^{th} data point at which a total of T_i fatalities was experienced.

t_e - termination time for fatalities clustered about line segment L_4.

t_L - time length of line segment L_4; thus, $t_L = t_e - t_s$.

t_s - starting time for fatalities clustered about line segment L_4.

$\bar{t}_j$ - arithmetic mean of time values in the j^{th} time interval; sometimes written as $\bar{t}_{L_j}$.

$\bar{\bar{t}}$ - the arithmetic mean of successive values of $\bar{t}$, where

$$\bar{t} = \frac{t_j + t_{j+1}}{2}$$

U - injury designation in the Medical Officer's Reports indicating that an aircrewman was lost, but it is not known how or where he was lost; the loss is thus unknown.

u - number of runs in a non-parametric run test.

u_0 - a positive integer such that

$$\sum_{u=0}^{u_0} f(u) \leqq \alpha$$

Var (u) - Variance of the parent probability density function from which the random sample of u-runs was extracted.

w - shorthand notation for $\ell n\ f^*(x)$, where

$$f^*(x_i) = f(x_i)/\sum_i f(x_i)$$

x - an independent real variable; may be discrete or continuous.

x_i - mid-point of the i^{th} frequency class interval.

- the i^{th} element in the set of all x's.

x_j - the j^{th} element in the set of all x's.

x(I) - computer flow diagram notation for x_i.

$x_{i,j}$ - element found at the intersection of the i^{th} row and j^{th} column in an Analysis of Variance matrix.

$\bar{x}$ - sample mean defined thus,

$$\bar{x} = \sum_{i=1}^{n} x_i/n.$$

y_i - observed frequency in the i^{th} frequency class interval.

y_i^* - normalized value of frequency in the i^{th} frequency class interval.

z_c - a standardized value, computed from given data, of an independent variable, assumed here to be normally distributed, and defined as follows:

$$z_c = [u - E(u)] / \sqrt{Var(u)};$$

it is sometimes designated as z_0. Here, c represents computed.

$z_{-\alpha}$ - a value of z such that

$$\int_{-\infty}^{z_{-\alpha}} f(z)\, dz = \alpha;$$

where f(z), here assumed to be continuous, is the parent probability density function of the real continuous variable z.

A.2 Greek Symbols

α - level of significance in the test of an hypothesis, equals the probability of a Type I error.

- a parameter in the gamma parent probability density function.

$\hat{\alpha}$ - an estimator of the parameter α in the gamma parent probability density function.

β - a parameter in the gamma parent probability density function.

$\hat{\beta}$ - an estimator of the parameter β in the gamma parent probability density function.

$\Gamma(\alpha)$ - the gamma function of the argument α, where $\alpha > 0$ represents a real number.

$\bar{\Gamma}(\alpha)$ - the arithmetic mean of two values of $\Gamma(\alpha)$.

$\Delta\bar{t}$ - time distance between successive values of $\bar{t}$.

μ_x - mean of the parent probability density function for the real variable x. A similar definition holds for μ_y.

σ_x - standard deviation of the parent probability density function for the real variable x. A similar definition holds for σ_y.

$\chi^2_{c/d.f.=4}$ - a computed value of the Chi-square statistic with four degrees of freedom.

$\chi^2_{T/d.f.=4}$ - the theoretical value of the Chi-square statistic with four degrees of freedom.

$\ell n\ y_i^*$ - natural logarithm of the normalized value of y_i, where

$$y_i^* = y_i \Big/ \sum_{i=1}^{k} y_i \;;$$

$\ell n\ y_i^*$ is sometimes written $\ell n\ f^*(x_i)$.

APPENDIX B

DEFINITIONS OF SELECTED STATISTICAL TERMS

APPENDIX B

DEFINITIONS OF SELECTED STATISTICAL TERMS

All definitions which follow are those for expressions contained herein. They are given to assist the reader as he studies this report.

Definitions

1. Class Interval--an interval within which persons, places or things having a prescribed quantified attribute can be found.

   Example: A study is to be made of a sample of 100 persons residing in Fairfax County, Virginia, having homes which range in value from \$50,000 to \$150,000. The study is to be sub-divided into housing value increments of \$10,000. The first class interval in the study consists of all persons in the 100 person sample having homes which vary in value from \$50,000 to \$59,999.99.

2. Confidence Coefficient--the confidence that can be ascribed to an attribute which a parameter in a given parent probability density function can have.

   Example: If $P(3 < \mu < 4) = 0.95$, then for a large number of random samples extracted from a given parent probability density function, the true parent mean will lie between 3 and 4 about 95 percent of the time. The interval $3 < \mu < 4$ is called the 95 percent confidence interval.

3. Continuous Probability Density Function--Define this probability density function by f(x). Then, f(x) represents the domain of a continuous real variable x, where the range of x is the infinite interval, $-\infty < x < +\infty$, a semi-infinite interval $0 \leq x < +\infty$, or a finite interval $a \leq x \leq b$. Further, f(x) is assumed to possess all other properties which make it a valid continuous probability density function.

   Examples of Continuous Probability Density Functions: The normal, t, F, Chi-square, gamma, exponential, and beta.

4. Discrete Probability Density Functions--Define this probability density function by $g(x)$. Then $g(x)$ represents the domain of a discrete real variable x, where x can only assume the discrete values $h, 2h, 3h, \ldots, nh$ for some real number h. Further, $g(x)$ is assumed to possess all other properties which make it a valid discrete probability density function.

   Examples of Discrete Probability Density Functions: The binomial, multinomial, Poisson, hypergeometric, negative binomial, and runs density function associated with a non-parametric run test.

5. Estimator--a number developed from a random sample which estimates a parameter in a parent probability density function. This is sometimes called estimate or sample statistic.

   Example: The sample mean, $\bar{x}$, is an estimator for the parent population mean, μ.

6. Frequency, Observed--the number of times which persons, places or things, having a prescribed quantified attribute, are observed to appear in a given class interval.

   Example: If ten persons own homes in the price range from $50,000 to $59,999.99 (see definition of Class Interval above), then the observed frequency in the first class interval is ten.

7. Frequency, Theoretical--the number of times which persons, places, or things, having a prescribed quantified attribute, would be expected to appear in a given class interval. This value is determined by evaluating the area, under the parent probability density function, over the class interval.

   Example: If the random sample of definition 1 above is extracted from a continuous parent probability density function, then, by definition:

$$n\,P(x_1) = F_1 = n \int_{50,000}^{59,999.99} f(x)\,dx,$$

where F_1 is the theoretical frequency in the first class interval. Thus, F_1 might not be a positive integer, as is required for the observed frequency.

8. Histogram--a bar graph or chart depicting observed frequency versus class interval for the set of class intervals which span the range of the data sample being studied.

9. Hypothesis--a generalized statement made about a parent population, the plausibility of which is to be tested (within confidence limits) by an analysis of a random sample, or analyses of random samples.

   Example: Refer to definition 1 above and formulate the following null hypothesis, denoted by H_0:

   H_0 : The average value of all homes in the parent population (see definition 1 above) is $80,000.

   An alternate hypothesis could be the following:

   H_1 : The average value of all homes in the parent population (see definition 1 above) is greater than $80,000.

10. Level of Significance--this is used to establish the confidence which can be associated with the rejection of a given hypothesis. This quantity is usually denoted by α.

    Example: If the null hypothesis, H_0, (see definition 9 above) is correct at the $\alpha = 0.05$ level of significance for a large number of random samples drawn from a parent probability density function, it can be expected, on the average, that 95 percent of these will have a mean value between $75,000 and $85,000.

11. Mean--this is a measure of central tendency either in a parent probability density function or in a random sample. The mean, μ, of a parent probability density function is estimated by the mean $\bar{x}$ of a random sample extracted from the parent density function.

12. Median--a measure of central tendency in a parent population or in a sample such that there are as many values above it as there are below.

Example: Given the random sample of size 7 as follows: 1.1, 2.2, 3.2, 4.6, 5.6, 6.5, 7.3, the number 4.6 is the median of this sample since there are three numbers in the sample less than 4.6, and three numbers greater than 4.6. As a matter of passing interest, the mean of the above sample is approximately 4.357.

13. Parameter--an unknown constant in a parent probability density function that can be estimated from an analysis of data contained in a random sample.

Example: If a parent probability density function is defined by the exponential equation $f(x) = be^{-bx}$, b is the unknown parameter in that equation which may be estimated from sample data.

14. Parent Probability Density Function--a mathematical function, discrete or continuous, which describes probability for specific values or ranges of values, of the designated independent variate.

Example: The normal probability density function, defined thus:

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

can be viewed as a parent probability density function.

15. Run--In a given random sample consisting of a dichotomous data set (such as 0's and 1's or a's and b's), a run is a sequence of elements of one kind followed by elements of the other kind, or followed by no elements at all.

Example: If a random sample consists of the following data set:

a, a, a, b, b, a, b, a, a b;

the set consists of six runs.

16. Standard Deviation--one measure of dispersion of data about the mean value. It is the square root of the variance.

17. Trend--an expression used in non-parametric trend analysis defined as follows: Given a data set $x_1, x_2, \ldots, x_i, x_{i+1}, \ldots, x_n$. Start with x_1, and compare it with all other elements in the data set. Count the number of times $x_1 > x_j$, $j = 2, 3, \ldots, n$. Repeat using $x_2, x_3, \ldots, x_{n-1}$. Get the sum of all such "trends" in the data set. An inference about an underlying trend in the data set can be made from the sum of all the inequalities $x_i > x_j$ for $i < j$, where $i = 1, 2, \ldots, n$.

18. Variance--a measure of dispersion about the mean. This can be a measure of dispersion about the mean of a known parent probability density function, in which case it is denoted by σ^2--the variance of the parent probability density function. It can also be a measure of dispersion about the mean of a random sample, in which case the sample variance is denoted by s^2.